1991

Influence Of Cost Consciousness And Heuristic Approach On Online Search Performance

Jill Austin

INFLUENCE OF COST CONSCIOUSNESS AND HEURISTIC APPROACH ON ONLINE SEARCH PERFORMANCE

by

Jill Austin

School of Library and Information Science

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
July 1991

ISBN 0-315-66310-3

ABSTRACT

This research used an experimental approach to investigate the relationship between cost consciousness, heuristic approach and recall in online searching. It also investigated some of the characteristics of the searcher that would be expected to influence heuristic approach, including familiarity with the search topic, familiarity with the database and anxiety. A 2x2 factorial design was used, where the factors were experience (expert/novice) and presence or absence of a cost consciousness treatment. Forty subjects in all searched the same two search questions. A verbal protocol technique was used, whereby subjects verbalized their thoughts while conducting their searches. The verbalizations were audiotaped, transcribed, and integrated with a transaction log to give a complete log of the terminal session.

Results indicate that the treatment was effective in eliciting feelings of cost consciousness and that searchers modified their behaviour in response to the treatment. The results further suggest that, under the conditions of this study, cost consciousness did not reduce either heuristic approach or recall attained. However, there is evidence that a greater heuristic approach resulted in higher recall. Two other findings of significance are that search performance was strongly influenced by the nature of

the search question and that an experimental approach that used written information requests had a strong negative impact on search performance.

The results suggest that heuristicity may be an important factor in search success. It appears that heuristic approach was influenced by the nature of the search question, but not by the experience level of the searcher. Cost consciousness did not have the effect hypothesized, that is, suppression of interactivity and information gathering behaviour. This result should be verified by conducting a similar experiment using more costly databases. There is evidence from this study that the low heuristic behaviour reported by other researchers may have been, in part, a result of the experimental designs used, which frequently employed a written search statement and denied the searcher interaction with the requestor. Finally, measurement of heuristicity proved problematic in this study, and it is recommended that further work be conducted to develop a good tool to measure this important phenomenon.

ACKNOWLEDGEMENTS

This project was supported, in part, by DIALOG Information Services Inc., who granted free online search time for a portion of the data collection. Their support is greatly appreciated.

The members of my dissertation committee brought diverse perspectives to the research and provided much appreciated support and constructive criticism. Mark Kinnucan and Roma Harris provided a solid theoretical base in cognitive psychology, though from differing perspectives. Jean Tague brought to the project a broad-based information science perspective and provided thoughtful commentary throughout. All three provided essential statistical and methodological guidance and advice.

I would particularly like to thank my advisor, Roma Harris, for her unfailing patience, advice, encouragement, and especially, sense of humour, throughout the years of work on this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

BACKGROUND

INTRODUCTION

One of the most interesting and challenging areas of research in the field of library and information science is that of searching bibliographic databases. Numerous studies have been conducted since the emergence of these databases more than two decades ago, all aimed primarily at maximizing the ability to extract from a database the information that exactly matches and answers an information need. There are several stages to this process: determining exactly what information might satisfy an information need, translating that need into a specific information request, identifying the sources most likely to contain the needed information, extracting the information from the database, presenting it to the requestor, and determining whether the information need has been met. The present research is intended to study aspects of the way in which searchers develop and execute a search strategy. Specifically, it will attempt to identify key factors influencing the success of the interaction between the searcher and the search system. Thus, a very small portion of the overall process has been selected for study.

There are three major questions which this research will address. First, what is the effect of cost consciousness on search success when searching an online database? Online searching is becoming increasingly expensive, with search costs incurred, depending on the vendor, by the minute, by the number of citations viewed and even by the number of search statements executed. Cost consciousness is believed to influence search behaviour, particularly the level of interactiveness with the search system, but the question of how it affects online behaviour and search success is unknown. The second question relates to the first. If a searcher is concerned about cost, it follows that the person will spend as little time as possible online, thus minimizing the overall interaction with the system. Yet the power of online searching lies in the ability to explore the database, try different strategies and learn a more effective approach to the problem through that interaction. The second question, then, is whether greater interactivity and a more heuristic problem solving approach will lead to improved search success, where heuristic problem solving refers to a trial-and-error approach that utilizes feedback from the system to improve the search. Finally, given that cost consciousness and interactiveness may be antagonistic to each other, how does the searcher respond to this conflict?

Various researchers have attempted to integrate knowledge about human cognition with knowledge of online

information retrieval. Blair (1980) referred to an 'anchoring phenomenon' from problem solving research, and both Smith (1976) and Fenichel (1980) described the search process as a problem reduction approach - again from problem solving research. Given that online searching is a problem solving task, cognitive research and its methodologies should offer insights to help in understanding the processes involved.

In treating online searching as a problem solving exercise, an analysis of what is known about the task itself, when combined with knowledge of what constitutes effective problem solving behaviour, should suggest key factors influencing the success of the search exercise. The section that follows addresses what is known about the task of online searching, and what are thought to be the most effective and successful behaviours for the task. The review focuses primarily on what has been learned about search behaviour in terms of individual differences, search experience, and cost consciousness, and what is known about the difficulties of obtaining high recall in searching. Following this is a review of theories of effective problem solving, drawn from the literature of cognitive psychology. The review of current theories of cognitive psychology focuses mainly on complex problem solving (that is, solving complex, multifaceted problems) and on individual differences in problem solving. Finally, these two fields of research are jointly analysed to produce a theory that

suggests which of the many possible factors in the online searching environment should be expected to significantly affect search success. The application of these theories to the specific task of online searching then leads to several questions, which in turn, generate the research hypotheses that were tested in this study.

REVIEW OF RESEARCH INTO ONLINE SEARCHING

Since the advent of online bibliographic databases, researchers and database developers have been trying to find ways to improve the interaction between the searcher and the system to achieve greater search success. Online searching is a difficult task. The searcher must learn the syntax of the query language for each online system, the searcher must know the content of each database to select the best databases for searching, the searcher must know the indexing and coding systems used for each database selected for search, and the searcher must design a strategy to apply to each database that will extract an acceptable number of citations (retrieving to the extent possible, only those citations that are relevant to the information need). In addition to these difficulties, the real world imposes budgetary constraints on most searchers, requiring that costs be kept to a minimum.

Given this rather complex process, questions are continually being asked about search behaviour: what

factors give rise to the best searching, what people make the best searchers and why, what are the most difficult aspects of a search and how can they be made easier, and how can we tell when a search is successful? Despite over 20 years of research, these questions remain largely unanswered.

What follows is a review of what has and has not been learned about online searching. This study is concerned only with the factors influencing the searcher, and not with database or interface design issues; that is, the research emphasizes a human problem solving approach to online searching. Thus, the body of research of most interest for the purposes of this study is concerned with individual differences in search behaviour, factors acting as constraints to problem solving, and aspects of the task that impose the greatest cognitive strain on the searcher.

Individual Differences

The study of human cognition has focussed on developing general theories about the skills involved in acquiring and using knowledge: skills such as memory, language, problem solving and decision making. It also involves the study of how individuals uniquely employ these skills. People may demonstrate similar memory processes, for instance; however, individuals will tend to differ in the way information about a specific task is stored.

One aspect of online searching that has been observed repeatedly is that there is tremendous variability among searchers performing the same search task. The interpretation of the information request will vary, the translation of the request into a search strategy will vary, the amount of interaction with the system will vary, and the assessment of the search result (degree of success) will vary. This would be true of any complex problem solving task, not just online searching. Given this variability, it is not surprising that a number of researchers have concentrated their efforts on seeking correlations between online search behaviour and various measures of individual differences.

The validity of the individual differences approach is exemplified by Fenichel's (1980) comprehensive review of research in online searching up to 1979, in which she noted that various researchers have found considerable individual differences in strategy formulation and search performance among experienced searchers. According to Fenichel "such large differences in search performance are consistent with findings of a similar and more extensively studied problem solving activity: computer programming" (p.117). She pointed out that research has indicated that groups of users can be distinguished on the basis of process variables, such as number of commands used, and that there is considerable variation in individual approaches to searching even when the same system and database are used.

In a study reported by Rowley and Butcher (1989), it was hypothesized that "different search intermediaries with similar training and using the same search tools and search parameters should obtain similar results when conducting an information search" (p.109). It would be expected that all searchers would identify a "core" of documents. The results of their study (in which subjects searched manually, rather than online) did not support this hypothesis. Rather, search results were inconsistent, overlap was quite low, and a consistent "core" of documents among searchers was not found.

Another example of a study that revealed large individual differences among searchers was done by Fidel (1987). She tested ten experienced searchers on two search requests, and then had them redo the searches two months later. She found large variation in search process variables across searchers, as well as differences for the same searchers over time. "This finding produced a new complication because it showed that the level of effort a searcher is using to perform an online search depends on situational variables, probably such factors as mood, additional commitments, or willingness to perform a particular search at a particular time. Therefore, if one is strongly committed to the discovery of the characteristics of the productive searcher, one should investigate situational variables" (p.58). She further cautions that studying the online search process in terms

of traditional search process variables (such as number of citations retrieved, number of search cycles, number of commands issued) is inadequate because those variables are strongly influenced by the situational variables. However, very little research has been done to investigate the impact of situational variables on the search process.

Several researchers have concentrated on seeking links between cognitive and personality traits and search performance (Bellardo, 1985; Woelfl, 1984; Logan and Woelfl, 1986; Saracevic et al., 1988; Logan, 1988). These studies generally involve the use of standard personality and cognitive tests to characterize searchers and look for correlates with their performance on one or more online search tasks. The results were largely inconclusive; however, there is some evidence that a preference for abstract over concrete learning style (as measured with the Learning Style Inventory Test) may lead to enhanced recall, as does a tendency toward enhanced verbal or language skills. In other words, search recall may be affected by aspects of the searcher's learning style, as well as verbal and language skills.

For example, Logan (1988) studied the relationship between learning modes and online behaviour for novice searchers. After administering three learning style tests to 76 novice searchers (Learning Style Inventory, Remote Associates Test, and Symbolic Reasoning Test), she found significant differences in online behaviour based upon the

groupings from the Learning Style Inventory. The main differences were between assimilators and accommodators. "Since assimilators are those who rate high in reflective observation and abstract conceptualization, and accommodators rate high in active experimentation and concrete experience, they occupy opposing quadrants in the Learning Style Grid" (Logan, 1990, p.509). "Novice searchers who rank in the Assimilation quadrant tend to spend longer online, enter more commands, complete more cycles, key more descriptors, and print more references during a search; those who rank in Accommodation quadrant tend to spend less time online, enter fewer commands, complete fewer cycles, key fewer descriptors, and print fewer references" (Logan, 1988 (p.205). Logan suggests that because novice searchers have little training and experience, they may rely more heavily on basic learning styles, whereas, experienced searchers may rely more on acquired techniques and less on learning style.

Another source of individual variability was described by Fidel (1985). She explored individual variability in searcher behaviour by studying ten experienced search intermediaries searching two test requests. By using verbal protocols, search transaction logs and answer sets in her analysis, she discovered that searchers exhibited a great deal of frustration in interpreting the requests without being able to interact with the requestor. The nature of the request was seen to affect search behaviour

and need to interact with the requestor. One form of variability described was the degree of specificity with which searchers interpreted the request. Fidel suggested that, to keep searcher variability to a minimum, researchers should keep test requests simple and straightforward. Thus, it seems that traditional research procedures, in which searchers are presented with a written information request, may enhance variability in search approach.

The knowledge of a subject that a searcher brings to the task represents another source of individual variability. For instance, Borgman (1986) explored this by comparing academic major with individual difference measures identified in studies of information retrieval and programming aptitude. Her findings suggest that academic major does correlate with ability to learn and use Boolean-logic based retrieval systems; for example, engineering majors show ability superior to that of arts and humanities majors. She suggested that a procedural model of thinking is required in the use of computers in general, and information retrieval systems in particular. "Automated systems require that problems be broken down into explicit facets with explicit relationships among them; much less structure is required for searching most manual systems" (p.21). Thus, it appears that the difficulty with using Boolean-based systems is somewhat reduced by the skill sets generally associated with engineering as a discipline.

Overall, then, the impact on searching of individual differences in certain cognitive characteristics has been reasonably well studied, especially vis-a-vis language ability, logical ability and learning style. These traits have been shown to have some effect on search performance and search success. In other words, there is some evidence that these factors contribute to individual variability in online searching performance, and the ability to quickly master a traditional Boolean-based retrieval system. At best, however, these studies have shown rather weak correlations with the search task.

One area of individual differences that has not been well addressed in the research to date is the effect on search performance of differences in problem solving style. One such variable is the extent to which searchers "trust" a system or place their faith in the system's ability to provide appropriate references (e.g., Harter, 1984b; Marshall, 1980). Some searchers may feel that it is necessary to interact heavily with the system to extract the information, whereas others might have the attitude that once the search terms from the initial strategy had been entered, the system would have to produce the best references. Thus expending a great deal more effort would yield little improvement. Since this type of variability in problem solving approach could greatly affect search performance, factors that affect interactiveness with the system are of interest in the present study.

Experience-based Differences

An important source of individual variability and a possible predictor of search success is the amount and type of previous experience that a searcher brings to the search task. Various studies have been conducted to examine differences in online search performance based upon varying amounts and types of experience. The most common focus of these studies has been to compare expert and novice searchers. This has helped elucidate the learning process for the task and aided in developing teaching methods for online searching. Indeed, in her review of information retrieval studies, Borgman (1986b) warned that it is critical in studies of searcher behaviour to take into account both experience and frequency of use of a system because they affect mechanical and conceptual problems in system use.

Searchers vary in the amount of general online experience they bring to the task, for example, the number of different databases they have used, and the amount of time spent searching. They can also vary in amount of experience they have had with a particular database. Fenichel (1981) provided a comprehensive analysis of searchers based on both of these types of experience differences. Her subjects varied in their general searching experience and in their experience with the ERIC

(Educational Resources Information Center) database. The purpose of Fenichel's research was to "discover those behaviors associated with the process of online bibliographic searching that are correlated with success" (p.23). Her method was to control as well as possible searcher and environment variables and to assess the impact of experience on search process and search outcome variables.

Some unexpected results were obtained from the study. Large individual differences in search behaviour were found within groups having the same experience level, as well as between groups having different levels of experience. The briefest and most cost effective searches were performed by the moderately experienced group - the group that contained the most subjects working in libraries that charge individual users a fee for searches based on connect time. "The only clearcut differences that could be attributed to experience were that the Novices searched more slowly and made more errors than the experienced subjects. However, there is some evidence that the searchers with the greatest overall experience who also had ERIC database experience had higher values on a group of measures called search effort variables" (p.29). Search effort variables include such things as the total number of commands issued, the total number of search terms used, the number of modifications to search terms, the number of citations viewed; in other words, the amount of effort or work the

searcher expended in trying to improve the search result. Fenichel found that subjects with ERIC experience used more thesaurus terms than did subjects without ERIC experience, and that in terms of recall and unit cost, the novices scored significantly lower, on average, than the experienced subjects. Nevertheless, the novices performed surprisingly well, and equalled experienced subjects in average precision.

A surprising result was that, in fact, a large number of subjects expended relatively little search effort. As Fenichel described it, "the simplicity of a large portion of the experienced subjects' searches was striking. In half of the searches the initial strategy was not modified and it was rare to find use of any but the most basic techniques of selecting and combining search terms" (p.30). She reasons that this may be a result of the straightforward nature of the experimental search topics and that most subjects retrieved a reasonable sized bibliography (between 3 and 50 references) in the first search cycle. Almost twice as many searches of the very experienced ERIC group were modified than were the searches of the other experience groups. "In more than two-thirds of the searches the subjects did not take the opportunity to review their retrieved references before giving the command to print all of them" (p.28).

Fenichel (1980) concluded her review of online searching with a summary of findings that included the

following: new users can learn to perform searches after a brief training period; there is considerable room for improvement in the searching level of many experienced searchers; for both experienced and inexperienced searchers the major problems were not with the mechanics of the system language but with search strategy; and that there is a substantial group of experienced searchers who perform rather simple searches, making little use of the interactive capability of the system. In other words, her results seemed to indicate that there is an initial hurdle that is relatively easily overcome in learning to use an online system, but that improvement beyond that initial capability seems to be more difficult to achieve.

Howard (1982) also looked at differences among searchers with different training and experience, and particularly differences among the searches of subjects with and without ERIC database experience. Her results and conclusions were similar to those of Fenichel. She used 42 searchers divided into five experience groups ranging from novices and moderately experienced searchers without ERIC experience to very experienced searchers with ERIC experience. There was evidence of greater search effort by very experienced searchers over moderately experienced searchers. ERIC experience was associated with greater use of thesaurus terms, while greater use of free-text terms was associated with lack of ERIC experience. Novices were slower but performed relatively well. Howard found that

"consciously or unconsciously the subjects tended to place more emphasis in their searching on precision rather than recall" (p.325). Highest recall was attained by the very experienced no-ERIC group, which may have been because they opted for a broader search strategy because of their unfamiliarity with the database. However, on the more difficult search, very experienced ERIC searchers had higher recall and precision scores and conducted the most cost-efficient searches. Howard reports a recall ratio for the more difficult search of only 9.64% for the total population. "This suggests that the level of experience and familiarity with the database may contribute to h er performance in a relatively difficult search" (p.321). The result of this study raises questions about what constitutes a difficult search, and about what aspects of experience enhance the searcher's ability to perform a difficult search.

The studies by Fenichel (1981) and Howard (1982) were conducted on the ERIC database. As a point of comparison, it is interesting to note the results of a study of end-users on a different database. End-users are non-professional searchers (not search intermediaries) searching in response to their own information need, and they are assumed to be relatively inexperienced, or novice, searchers. Sewell and Teitelbaum (1986) analyzed end-user behaviour on MEDLINE (a database produced by the National Library of Medicine) over an 11-year period. They

concluded that end-users performed simple searches, mostly using the Boolean operator AND, neglecting the OR and NOT operators. They failed to use the most powerful features of the system, subheadings and explosions, leading in a few cases to failure to retrieve a substantial number of references relevant to their request. These results substantiate those of Fenichel and Howard, suggesting that lack of search experience seems to correlate with a more simplistic problem solving approach.

Experience-based differences were further analyzed by Oldroyd (1984), who compared the search strategies of experienced and inexperienced searchers and found that experienced searchers were better able to identify search terms appropriate to the nature of the query and to the structure of the file being searched. They were also better able to identify synonyms, including acronyms and subject codes, than were the inexperienced searchers:

> A search carried out by the individual user, after a short training session, almost invariably results in his finding some relevant documents fairly easily. However, he is certainly lacking in experience and technique when it comes to achieving high recall. He is even more likely to fail to find relevant documents when it becomes necessary to change strategies if a first search statement is not successful. Naturally, in most cases, this basic search will not retrieve more than a part of the total relevant documents available in the database. (p.233)

This research may shed some light on the previously described aspect of experience, wherein the novice

overcomes the initial hurdle of using the system fairly easily, but has difficulty progressing to more sophisticated search behaviour. This study seems to indicate that one of the barriers is lack of familiarity with search terms and that a second barrier is the searcher's inability to substantially revise an ineffective strategy.

Also revealing is the work of Fidel (1984; cited in Fidel, 1987), who found "that experienced searchers consistently used certain databases to help them to formulate a search strategy, and that they were willing to follow leads that were suggested by the retrieved citations, even though these approaches deviated from concepts included in the original request" (p.59). She noted that novices may be more reluctant to follow new leads in the course of the search.

Another difference in search behaviour between novices and experts was described by Harris (1986). She compared experienced and novice searchers in terms of sequences of moves made between search cycles, using transition analysis techniques, and found that for novices, the size of the retrieved set was an important factor in determining whether or not searches continued at the end of a cycle. This relationship was not found for experienced searchers. Professional searchers were found to have more complex search formulations and to display greater search effort, in terms of using more SELECT and COMBINE commands to

achieve results. However, some caution must be used in generalizing from these conclusions given that different questions were searched by the two groups and that the experienced searchers were conducting 'real' searches while the novices were searching in an 'artificial' setting from a written search question.

An interesting study by Marchionini (1989) looked at the behaviour of children searching a full-text electronic encyclopedia. He observed 28 third and fourth graders and 24 sixth graders. The older children favoured moves examining the title and text, whereas the younger children generally favoured query refining moves. Nine searchers started with initial hits and subsequently lost them during search reformulation. Often this was a failure to extract the relevant information when the appropriate text was, in fact, retrieved. Marchionini questioned whether this occurred because the children lost sight of their search goal because they were focussed on the system, the situation or on reading the text.

To summarize, it appears that relative to more experienced searchers, novices, although searching surprisingly well, tend to score lower on search effort variables, have lower search success in terms of recall and precision, and not surprisingly, make more errors. Experienced searchers, on the other hand, seem to have greater facility in selecting search terms appropriate to the query and the database, perform more complex searches

(possibly as a result of greater facility with the command language and greater knowledge of the capabilities of the search system), and, like novices, tend to have problems formulating search strategy.

It is still unclear which types of experience are most significant in altering search performance, given the number of experience variables: experience with retrieval systems in general, experience with specific retrieval systems, subject experience, database experience, and frequency of use of various systems and databases. Thus, it would be useful to know how experienced searchers develop their superior capabilities in selecting search terms (and synonyms) appropriate to the query and database. Finally, it is not known why expert searchers often exhibit very low interactive searching, resembling the search performance reported for novices and end-users. It is possible that this is related to a combination of learning style and experience. It may also be a function of experimental design and situational variables.

Given the strong evidence that experience level significantly influences search behaviour, it is essential that searchers' experience level be controlled for in any study of the search process. Thus, in the present study, in which cost consciousness and heuristic search approach are investigated for their effect on search performance, experience level is included in the experimental design. In addition, the consistent findings of previous

researchers that novices tend to score significantly lower on recall than do experts suggests that recall is a useful measure for differentiating between searchers in terms of search performance. Because of this, recall was selected as a specific search goal for subjects in the present study.

Cost Consciousness

Online searching can be an expensive form of information gathering. Every year online searchers are faced with increased charges with each logon command, each search of a database, each command typed and each reference viewed. At the same time, there are more people wanting access to ever-increasing numbers of databases. With shrinking library budgets, it is inevitable that cost should have an effect on searching. Indeed, cost consciousness, as described in the following review, is frequently proposed as an important factor influencing the search process.

In her 1980 review, Penichel (1980) noted several studies pointing to cost as a motivation to keep searches simple and to minimize terminal interactions. She suggested that cost was a factor in the behaviour of a substantial number of her subjects, and that institutional setting probably influenced both cost consciousness and search formulation style. "Different user groups can have

different recall/precision requirements that can influence how a search is performed. The various methods of paying for searches can also affect performance. For example, it is reasonable to expect that if cost pressures are great, as when an individual requestor is paying personally, the searcher is likely to execute a simpler search than he or she would perform if cost were relatively unimportant" (p.113). She concluded that the search process is sensitive to many variables less obvious than the skill of the searcher and nature of the question; these include management policies and charging procedures.

As Bates (1981) pointed out in her review of search techniques, "the heart of a chapter on search techniques should be the section that describes research done to compare different strategies for their cost and productivity under different circumstances. Not a single study doing just that was found" (p.150).

Fidel and Soergel (1983) noted that searcher attributes have been widely described in the literature, and range from personality attributes to detailed analysis of online experience. "Of the attributes that are easily defined, only a few have been proven to have any effect on online bibliographic retrieval. Of all searcher characteristics, cost-consciousness is regarded by most researchers as a major factor affecting online searching" (p.166).

Fidel (1983) looked at the effects of user charges on online searching behaviour and concluded that while there is some evidence that searchers in fee-charging settings perform more efficiently, there is no real evidence that searching behaviour is affected. The impact of cost on productivity is difficult to measure because it varies with the factors selected to measure the productivity. She pointed out that the effect may be altered with the procedure used for charging users. Searchers tend to develop a particular searching behaviour within a particular setting and that behaviour may carry over to other settings, including experimental settings.

Morris et al. (1989) interviewed search intermediaries to identify search techniques that might be incorporated into an expert system to assist in search modification (MOSS). They found an iterative approach that consisted of searching, logging off, reformulating, and redoing the search. "The high costs of online searching, particularly connect time charges, are the most common reason for this type of searching and are a considerable deterrent to the near-instantaneous interaction advocated by the textbooks" (p.425). They also concluded that "worry about connect time charges being incurred certainly results in little modification during an online session and an unwillingness to experiment" (p.426).

Byrne (1989), using a mailed questionnaire, examined a number of characteristics of online searchers in Australia.

He received 238 responses (84.5%) from a random sample of searchers from various work settings. Half the respondents performed searches in the absence of the requestor. In terms of perceived satisfaction of clientele with searches, he found that "there appeared to be some under-confidence among those charging and some over-confidence among those offering free services" (p.401). Those offering free searches were more confident of search success in databases having controlled vocabulary, and tended towards greater interactivity while searching. "Charging tended to engender cautiousness and the choice of approaches designed to minimise cost" (p.401). The most cost-conscious were significantly more likely to prepare alternative search strategies. He found that most respondents to his survey rejected the fast-batch search approach. "Amount of interaction often depended on the circumstances, including the purpose of the search, the concern about cost and the organisational environment" (p.404).

In other words, it is suspected that cost concerns may be related to low interactive searching, and that certain environments, particularly fee-charging settings, may affect search behaviour. However, no studies have attempted to directly manipulate cost consciousness to determine its effect on search performance. Another purpose of the present study, then, is to use an experimental approach to examine the influence of cost consciousness on search behaviour, particularly level of

interactivity with the search system, and to assess the effect of cost consciousness on search success.

Recall and Precision

Recall and precision have been the most common measures of search success, where recall is the proportion of relevant retrieved citations to the total relevant references in the database, and precision is the proportion of relevant references retrieved to the total retrieved. It has often been stated that there is an inverse relationship between recall and precision, such that one is enhanced to the detriment of the other. There are certainly problems with recall and precision as measures of search effectiveness; however, having nothing better to replace them with, they continue to be the preferred measures of search success. Many attempts have been made to identify factors in online searching that affect recall and precision. The results of several of these studies are described below.

Several researchers have reported low recall levels in studies of search performance (e.g., Lancaster, 1972; Fenichel, 1981; Saracevic et al., 1988). Recall values in these studies were reported at 22% to 58%. Precision values tended to be higher: average values reported were from 63% to 86%. It is not surprising that high recall is

more difficult to obtain than is high precision. To increase precision, the searcher must remove non-relevant references from an answer set that is often small enough that a substantial proportion of citations can be browsed for relevance determinations. To increase recall, the searcher must extract relevant references from the database to add to the answer set, without knowing how many relevant references exist.

Saracevic and Kantor (1988a) investigated the effect of various properties of the search question, such as subject domain, clarity, specificity, complexity and presupposition (presence of concepts implied but not specifically stated), on recall and precision. Specificity of the subject and complexity of the question were the only characteristics significantly related to precision. They found that "questions with low specificity and high complexity have twice the odds that precision of searches be high" (p.188). None of the question characteristics were significantly related to recall scores.

Saracevic and Kantor (1988b) looked at the effects of tactics and efficiency measures on recall and precision. These were measures of number of commands, command cycles, and search terms, along with preparation time, online connect time, and total time used. Not one of the measures of search tactics and efficiency was found to significantly affect precision or recall. However, they did find a significant difference in recall for searches based on

different sources for the query statement (although no difference in precision was found). "The best recall was from a search type done on the basis of a taped problem and intent statement by the users. The poorest recall was achieved when words from written questions were used as search terms without any elaboration (as if they were picked automatically)" (p.211).

Another study that investigated factors influencing recall was that done by Harter (1990), investigating search strategy and postings overlap with respect to online retrieval. He found that "because overlap among relevant postings in elementary posting sets is slight, a search achieved by combining only a few search terms, among many more reasonable candidates, (sometimes called a briefsearch, a fast batch search, or a quick-and-dirty search) cannot achieve high recall" (p.144). Elementary postings sets were defined to be the result of the intersection of a single term from each facet in a building blocks search strategy. These elementary postings sets were characterized by poor recall and precision, and were found to often contain a great many terms that were present in relationships other than that intended by the searcher. To enhance search performance, it was suggested that feedback mechanisms should help searchers to identify the elementary postings sets contributing to the poor retrieval so that these sets can be dropped out of the intersection of sets, or so that other terms could be added to reduce

the ambiguity of the terms. Harter concludes, "the present study suggests that searchers should recognize that any particular recall or precision goal can be achieved in a given search. The main point, then, is to *have* a goal, and to work in a rational way toward achieving it" (p.144).

In summary, it appears that high recall is difficult to achieve for even the most experienced searchers, whereas high precision can be achieved by novices. The type of experience affecting recall is not entirely clear, although it seems that high overall searching experience is more significant than high database-specific experience. The extensive study by Saracevic et al. (1988) has shed more light on factors affecting recall and precision. In terms of the searcher, skills in word association and preference for abstract learning were related to higher search success. Of the question characteristics investigated, specificity and complexity affected precision, but no such characteristics seemed related to recall. Receiving more contextual information about the search question seemed to lead to improved recall. Overall, it seems that the factors influencing recall remain elusive, but factors influencing precision are more clearly defined.

It is not known how the interaction with the system during the search affects the success of the search, nor how variables related to that interaction and specific to the individual for a particular question in a particular search setting affect recall and precision - variables

such as anxiety, familiarity with the topic, and cost consciousness. One purpose of the present research is to address these issues.

Problems with Strategy

Search strategy refers to the analysis of the search topic into the concepts involved, and the translation of those concepts into the language of the online database. This involves finding search terms representing the different concepts comprising the topic and coordinating those terms using Boolean logic. Appropriate search strategy is the key to the success of a search. However, formulating that strategy is problematic for many searchers. As a result, a great deal of research in online retrieval has focussed on searchers' problems with formulating and executing search strategies. These investigations have examined, among other things, the aspects of strategy formulation and execution that produce the greatest levels of cognitive strain, as well as techniques for reducing this strain. Cognitive strain refers to excessive mental stress or pressure, which may reduce the effectiveness of an individual in performing cognitive tasks.

Standera (1978; cited in Bates, 1981) investigated the psychology of online searching by questioning experienced searchers. He identified 17 phases in the

online search process and the pressure points along it. "The points of highest pressure, interestingly, were strategy design and modification" (p.150).

This was reiterated by Wanger et al. (1976; cited in Borgman, 1986b), who carried out a survey of online searchers in which it was determined that respondents reported difficulty in formulating search strategies "some" (47%) or "most" (8%) of the time, and that 36% reported difficulty in making relevance judgements some of the time. Further, Borgman noted that "searchers often fail to consider the inverse relationship between recall and precision in searching, not recognizing that it is necessary to accept a low score on one measure to achieve a higher score on the other" (p.389).

Several studies have looked at sources of cognitive strain in conducting an online search. Vigil (1983) proposed an algorithmic method of searching that would reduce cognitive strain that arises from demands that the interface to a Boolean retrieval system place upon human memory and human information processing capabilities. He suggested that a great deal of cognitive strain arises from the difficulty of comparing and contrasting multiple retrieved sets, and proposed using Boolean negation to eliminate redundancy in retrieved sets. In this way, the searcher would evaluate the effectiveness of each modification in strategy by ascertaining how much overlap new sets had with earlier sets. He used an experimental

setup with ten novice searchers divided into two groups of five (Vigil, 1982). One group received training in the use of the Boolean operator "NOT" to eliminate redundant citations in retrieved sets. The other group was a control. Vigil found that the experimental group's performance on all effectiveness scores was 30% higher than the control group's. It appears, then, that searchers' effectiveness can be improved through learned tactics; however, it is not clear that cognitive strain is reduced by using these tactics.

Another serious problem in conducting an effective search is in coping with large retrieved sets. Blair (1980) investigated searching biases inherent in interactions with very large databases. These biases are a result of information processing strategies by which searchers cope with large retrieved sets. Searchers must predict, by means of a formal search query, the terms used to index the documents of interest; however, it is not enough to predict the correct terms. Instead, the number of documents retrieved must also satisfy what Blair terms the 'futility point criterion.' "This point is the maximum number of retrieved documents that an inquirer would be willing to begin browsing through. It represents the largest size retrieved set of documents he is willing to look at" (p.271). Thus, success in searching is dependent upon satisfaction of both the prediction criterion and the futility point criterion (FPC).

The strategy often used by searchers to cope with this difficulty is to reduce the retrieved set by combining with another term. Very often the searcher exhibits an 'anchoring phenomenon' (from the literature of psychology, this is a strategy for assessing the likelihood of uncertain events). Judgements of unknown values are excessively influenced by the initial value obtained or starting point. In online searching, searchers cling to the set retrieved by the first term and add successive terms to it using Boolean AND in an effort to quickly reduce the size of the set to below the FPC. These successive terms will be reevaluated and replaced as necessary, but the initial term often remains unchanged. "By keeping the anchor set intact (or by being willing to change it only as a last resort) the inquirer is, in effect, overestimating the probability of the conjunctive event that all the terms in the anchor set will be assigned to the document(s) he wants" (p.275). It is not known to what extent searchers exhibit this phenomenon, nor to what extent it influences search success.

The results of these studies suggest that strategy formulation may be the most difficult task for the searcher, who must satisfy several criteria, including predicting the best search terms and retrieving sets that satisfy an internal set size, or futility point, criterion. One indication of the degree of difficulty this poses for searchers is that there is surprisingly little

intersearcher consistency in formulation of search strategy for the same query.

Given the importance of strategy formulation, it would be useful to learn more about those aspects which are most difficult for the searcher. It would also be useful to know how a particular strategy is chosen for a query, and to what extent the searcher adheres to that strategy during the interaction or is willing to alter strategy with feedback from the system. In other words, how does information obtained in the course of the search affect strategy and the interpretation of the query, and to what extent is the searcher open to changes in strategy in response to system feedback? In the present research, the effect of the amount of interaction between the searcher and the system on search success will be examined, and the effects of certain searcher characteristics on the overall interactiveness exhibited in the course of the search will be tested.

Cognition and Online Searching

Various aspects of cognitive research have provided a theoretical framework for the investigation of information retrieval, including categorization of objects, storage and retrieval of information from long- and short-term memory, development of mental models, and decision making. A good analysis of the relevance of these theories to

understanding online search performance is provided by Fenichel (1980):

> To this reviewer, the most promising concepts for providing a frame of reference from which to understand the search process are those having to do with human information processing. Specifically, it is easy to see an analogy between a person as a serial processor of information, integrating several 'percepts' into an overall percept and then discarding the component percepts, and the type of online search in which the searcher deals with one concept at a time and then relates all the individual concepts to one another. The notion that there is possibly a limit to short-term memory also seems relevant, as do the findings that human performance degrades when information workload is too small or too great or when task complexity is too low or too high. (p.108)

The following is a review of studies undertaken to investigate online searching within a cognitive framework. Investigations into types of information retrieval other than online retrieval have provided insights that can be applied usefully to online searching. For instance, Ingwersen et al. (1980) studied the reference process as a problem solving task and used 'thinking aloud' techniques to study cognitive processes involved in reference interactions for reference librarians having several years of experience. They found that "procedures were begun without reflecting upon topic-to-topic relations. This is in accordance with the training in reference work which by and large is 'title training' and classification saying 'that a book is classed in one place in the system'. Training in librarianship may differ from place to place.

However, experience from the use of computerized retrieval systems elsewhere seems to indicate a similar lack of subject analysis" (p.168). The results of this study are important to the current research in two respects: (1) it is unlikely that manual and online reference procedures will differ radically in approach to subject, and (2) the methodology using verbal protocols was found to be an effective technique for evaluating and interpreting the search process.

The study of mental models has been applied with some success to online searching. When people interact with a new system, such as a computerized retrieval system, they develop a picture or model to explain how the system works, and the model allows them to predict the response of the system to a particular action they might take. The more accurate the model, the more successful they can be in their predictions. Thus, they can gain competence with the system faster with a good model than with a poor one. Borgman (1983) tested the hypothesis that people can be trained to develop a mental model of a system which will aid in solving complex tasks. She applied this theory to searching online catalogs by training two groups of subjects with either procedural instructions or conceptual instructions and observing their performance on searching tasks. She found that on simple searches the two groups performed equivalently, whereas on complex searches, the conceptually trained subjects had superior performance;

however, the process by which they performed better was not clear.

Allen (1990) used an experimental approach to evaluate the way information system users organize their knowledge of a topic, and its effect on the information interaction. He had subjects read one of three texts on a search topic and make notes in one of three ways: assuming they would be requesting an online search, categorizing their notes in terms of structural elements of the text (e.g., author, methods, findings), and given no note-taking instructions. He then looked at the manner in which they filled out a simulated pre-search request form. Except for one question, all noticeable differences in the notes taken by participants disappeared in their responses to the online search form questions. "One possible explanation of this phenomenon is that the cognitive structures imposed by the note-taking were superseded by the structures implicit in the questions on the pre-search forms" (p.540). He found that subjects who were more familiar with information retrieval functions responded to open questions with short answers and fewer idea units. "This research indicated that the user's organization of knowledge about the search topic can be a useful input in selecting the type of questions to be posed by the intermediary in some topic areas" (p.540). The study in interesting in its attempt to evaluate the pre-search input that search intermediaries can elicit in formulating search strategy and has

implications for the study of the effect of users mental models of the functions of an information retrieval system.

Another area of cognitive research that has been usefully applied to online searching is that of decision making. Problem solving involves making a series of decisions about, for instance, the best path to follow, when it is time to change course, and recognizing when a solution has been found. Blackshaw and Fischhoff (1988) investigated decision making during searching of an online public library catalog by 60 subjects. Subjects were tape recorded during their search tasks and were asked "to describe their goals, their favored alternatives, and their confidence levels (in quantitative terms), as well as whatever else they chose to relate" (p.380). They found that the subjects' "performance resembles that revealed in studies of decision making in other contexts. In particular, people are only moderately sensitive to the likelihood of their succeeding, being overconfident for all but the easiest of tasks" (p.369).

Heuristic Approaches to Problem Solving

Of all the cognitive theories applied to the study of online searching, the study of heuristics has been, perhaps, the most illuminating. Fidel (1986, 1988) analysed the heuristic knowledge held by expert searchers. Heuristics, in these studies, referred to the "rules-of-

thumb" that people use to solve a problem. They are the devices gained through experience that usually offer a degree of success in problem solving. Heuristics generally accrue with experience. Fidel used verbal protocols and interviews to elicit from eight experienced searchers heuristics on selection of search keys. She was able to reveal a routine for the selection of search keys (free text or controlled) along a decision tree. The selection of a search term involves two questions: can the term be mapped to a descriptor, and is the term 'good' for free-text retrieval. "The selection routine is not deterministic; it cannot always accurately predict the selection of search keys unless other factors and their impact are known" (1986, p.38). The purpose of the research was to determine whether expert system intermediaries could be developed for online searching. The results seemed to indicate that an explicit knowledge representation is held by experts in the domain of online searching, and that the organization of this knowledge can be elicited and described.

One important aspect of the study of problem solving has been to look at deficiencies in problem solving approach. Harter (1984a) took this approach in a survey of Florida searchers, from which he reported that when searchers were asked to select from a list of activities those that were most like online searching, some of which were deliberately algorithmic rather than heuristic in

nature, 9 out of 72 respondents chose the algorithmic activities. These included looking up a number in a telephone book, using an automated bank teller, and using a calculator. Nearly half of the respondents indicated "that trial-and-error methods reflected fuzzy thinking and poor search preparation, equating trial-and-error methods with 'fooling around' online" (p.255). He found that academic librarians were more cost conscious and tended to "consider cost to be the most important factor in deciding when to terminate a search" (p.255). Several other measures indicated a tendency for academic librarians to view the heuristic approach to searching with less favour than a more algorithmic approach. He also reports that with experience comes flexibility - a willingness to 'play' it by ear,' to adapt to changing conditions, and to interact at the terminal.

Bates (1989) suggests that an information request is not satisfied by a final retrieved set, but is satisfied "by a series of selections of individual references and bits of information at each stage of the ever-modifying search" (p.410). She refers to this sort of approach as berry-picking. This model assumes that the search query itself is shifted and modified as the search proceeds, which fits well with the view of the online search as a heuristic process.

Vigil (1983) also looked at deficiencies in problem solving behaviour of searchers, and pointed out that "most

online searchers are not interacting with the computer commensurate with the capability available" (p.281). The greatest strength of the online system is "the ability to use feedback, iteration, and the heuristic problem solving process, i.e., interaction. Yet it seems that this may not be the mode of searching used by most searchers" (p.281). He proposed that it is the demand for simultaneous cognitive processing in comparing and contrasting various sets generated that places considerable strain upon the searcher. Bringing together the parts of the search into a cohesive whole is critical, and "may be the limiting factor in the ability to search interactively" (p.284).

Wanger et al. (1980) conducted an extensive study of searchers of the National Library of Medicine databases. Using real requests submitted by searchers, they assessed the searcher's online behaviour and performance. An expert panel performed relevance judgements on search results, so that recall and precision could be calculated. They also used a background questionnaire to gather information about the searchers' typical search behaviour and their procedures used in planning and conducting searches. They found a surprising lack of interactivity in many of the searches performed, and coined the term 'fast batch' searchers to describe this behaviour.

Hawkins and Wagers (1982) describe an interesting search behaviour, which they called "interactive scanning." They describe this as a "search technique which makes the

best use of the interactive qualities of online systems and the feedback available from careful heuristic searching" (p.13). The searcher starts by formulating a broad concept to capture most of the literature on the topic, retrieving a high postings set. A number of citations are then viewed - as many as necessary to provide a picture of the literature on the topic. The strategy is then reformulated in a series of modifications, using an iterative approach. They describe the technique as "especially valuable in those cases in which the topic is foreign to the searcher and the user requires high recall" (p.13). It is useful to keep this technique in mind in evaluating behaviour of searchers in experimental settings, using written questions, in that they may be in a situation where they have little background information to help in formulating a strategy, and must, therefore, rely on the information system to give them feedback about the topic. In other words, this provides a model for a type of heuristic search behaviour.

Harter (1984b) proposed that searchers should adopt a model of scientific inquiry for online searching. He indicated that problem solving attitudes and abilities are the most important characteristic in determining searcher success. Results from past research suggest an "excessive rigidity in search preparation and execution for some searchers, an overly algorithmic approach to online searching" (p.115). Harter speculated on the results of

various studies (e.g., Wanger et al., 1980; Fenichel, 1981) in which more than half of the searchers studied did no intermediate printing or browsing. He suggested that these 'fast batch' searchers may be influenced by time and cost factors, or may believe it is not their responsibility to evaluate the search results; "or that they feel that the presence of the terms requested in a citation means that the citation must therefore be relevant to the request; or that they may simply have 'blind faith' in the vocabulary and the system" (p.115). They appeared to be acting deterministically, rather than probabilistically. If searchers were to use the model of scientific inquiry, they would formulate hypotheses about the search problem, operationalize definitions of 'concept' variables by selection of vocabulary and search fields, and evaluate results of tests of the hypotheses by browsing and selecting from retrieved sets. Instead, however, it appears that they often fail to evaluate the results of the tests of their hypotheses.

An anecdotal report by Marshall (1980) reiterates the concern over lack of interactivity with the search system by experienced searchers. She reported comments at an online meeting to the effect that only the end user would be able to make a value judgement about the results, so that it was a waste of time for the searcher to check the results. Another searcher reported sending all search results directly to the end user without evaluating the

results. Marshall referred to these searchers as "non-evaluators" and expressed concern that they might be representative of online searchers.

In the various studies that have attempted to link what is known about human cognition (especially problem solving, decision making, and mental models) to the task of online searching a number of useful findings have emerged. First, it appears that the use of verbal protocols, or thinking aloud techniques, provides an effective tool for understanding the processes involved in the online interaction. Second, the idea that searchers formulate mental models to assist them in their interaction with online systems shows promise and suggests that if differences in the mental models held by experienced and novice searchers could be identified, those resulting in greater search success could form the basis for more effective training in search techniques. Finally, the studies have clearly revealed that many searchers take an overly algorithmic approach to their task. However, the reason for this remains unclear. Cost consciousness has been suggested as one reason for low interactivity as well as failure to use the most powerful features of online retrieval. Thus, one purpose of the present research is to investigate some of the factors hypothesized to lead to a more algorithmic rather than heuristic search approach, and the effect of this approach on search success.

The various studies described above have all dealt with the factor referred to as heuristicity, heuristic behaviour or heuristic approach. Various aspects of heuristicity in online searching have been described. These include being flexible, adapting to changing conditions, being interactive at the terminal, and responding to feedback from the system. All of these aspects have one thing in common - there is no clearcut method by which they can be measured. Measuring such things as flexibility and adaptability is problematic at best. Similarly, measuring responsiveness to feedback from the system is difficult. Interactivity at the terminal is somewhat easier to measure, in that the traditional measures of search process, such as number of search cycles and number of citations viewed, give an indication of the amount of interaction taking place. Nevertheless, it is clear that defining a measure for heuristic approach is a difficult task.

One method of measuring heuristic approach might be to have coders assign to a particular online search session a rating for heuristicity. The possible sources of data for this coding would be the transaction log capturing the keyboard input and the system responses, a transcript of the searcher verbalizing his or her thoughts while searching, and any notes that the searcher makes before and during the terminal session. Another method of measuring heuristic approach might be to derive an index value for

heuristic approach based upon several measures of the search process that individually capture some of the aspects of heuristicity. These would include measures that suggest interactivity (such as number of commands issued), as well as indicators of utilization of feedback obtained during the search (such as addition of search terms identified during the search as being potentially valuable). One of the goals of the present study is to develop a measure for heuristic approach exhibited by a searcher.

Summary

It is apparent from this review of research that online searching is a complex process. There are large individual differences in approach to the task, yet it is not clear how those differences affect search success. Clearly, experience is a factor in search behaviour and search success, but the nature of that effect is not clear, and even inexperienced searchers have been found repeatedly to perform surprisingly well. Concern over search costs is widely viewed as a significant variable affecting the search process, yet no studies have attempted to empirically test its effect. Some factors affecting search success (in the form of recall and precision) have been studied extensively. It is clear that high recall is difficult to obtain. However, information on factors

successful in increasing recall is sorely lacking. Various researchers have commented on the low level of interactivity of even experienced searchers, but it is not clear what factors influence this nor how it affects the overall success of the search. Theories and methods borrowed from cognitive psychology have been shown to be useful in elucidating the search process. The view that within the theory of human problem solving lies an explanation of the sometimes counter-intuitive results that various researchers have discovered in the behaviour of online searchers will be further developed in the next section.

PROBLEM SOLVING

This section will describe theories of how people solve problems, starting with a general theory of human information processing, a description of research into solving relatively simple problems, and a theory of decision making. Refinements to these theories are then described that take into account approaches to solving more complex problems and individual differences in cognitive style. Throughout this review, these theories will be related to what is known about online searching.

General Theory of Problem Solving

There are two general methods of approach to solving a problem: one is a heuristic approach, the other is an algorithmic approach. An algorithm is a step-by-step approach that guarantees success when applied to a particular type of problem (Reed, 1982), an example being calculation of income tax or mortgage payments. In contrast, heuristics are often successful in solving a problem, but do not guarantee success. A heuristic approach involves trial-and-error attempts at problem solving, where the problem solver continually gathers information by testing different approaches to the problem, applying 'rule-of-thumb' or 'tried-and-true' methods. Examples are means-end analysis, forming sub-goals, working backwards and forming analogies. Means-end analysis involves comparing intermediate problem states (the state following a problem solving move) with the final goal state. If a move results in an intermediate state that looks more like the final goal state, than that move is considered successful. In forming sub-goals, the problem solver identifies a sequence of stages (or sub-goals) in solving a problem and concentrates on attaining the sub-goals, as opposed to concentrating always on the final goal or end state. In working backwards, the problem solver starts with the desired final goal state and works backward to the initial problem state. A good example of this would

be solving a mathematical problem by working backwards from the answer to the problem. Finally, problem solving by analogy involves identifying a similar problem for which the solution is known, and modifying the solution to fit the current problem. For instance, in trying to find an item in an online catalogue, the problem solver might assume the online catalogue is structured like a card catalogue and would apply the familiar methods of searching by title, author or subject.

Heuristics have been studied predominantly using games and puzzles, which present well-defined problems to the solver and permit straightforward interpretations of behaviour. The types of problem studied fall into three general categories: problems of arrangement (e.g., anagrams), problems of inducing structure (e.g., analogy problems), and transformation problems, which are distinct in providing the solver with a goal state to be attained. An example of a transformational problem is the Tower of Hanoi problem, where the problem solver is presented with three upright posts, each having three rings arranged in increasing diameter from top to bottom. The problem solver is told to change the arrangement so that the nine rings are arranged on each post in decreasing diameter from top to bottom, with the constraint that only one ring at a time can be moved.

Much of the problem solving research has been stimulated by workers in artificial intelligence and has

been aimed at programming a computer to solve problems, on the assumption that this would allow a model to be derived for human problem solving. The work of Newell and Simon emphasizes this goal to the extent that they evolved a theory of human problem solving in which the person is seen as an information processing system acting on inputs and producing outputs that can be fed back into the processor.

A general organization of the problem solving process is given by Newell and Simon (1972, p.88):

1. An initial process, here called the input translation, produces inside the problem solver an internal representation of the external environment, at the same time selecting a problem space. The problem solving then proceeds in the framework of the internal representation thus produced - a representation that may render problem solutions obvious, obscure or perhaps unattainable.

2. Once a problem is represented internally, the system responds by selecting a particular problem solving method. A method is a process that bears some rational relation to attaining a problem solution, as formulated and seen in terms of the internal representation.

3. The selected method is applied: which is to say, it comes to control the behavior, both internal and external, of the problem solver. At any moment, as the outcome either of processes incorporated in the method itself or of more general processes that monitor its application, the execution of the method may be halted.

4. When a method is terminated, three options are open to the problem solver: (a) another method may be attempted, (b) a different internal representation may be selected and the problem reformulated, or (c) the attempt to solve the problem may be abandoned.

5. During its operation, a method may produce new problems - i.e., sub-goals - and the problem solver may elect to attempt one of these. The problem solver may also have the option of setting aside a new subgoal, continuing instead with another branch of the original method.

This description is completed by adding that "the continuous influx of new information from the external environment may offer new solution possibilities or demands that cause the problem solver to interrupt its current activities to try different ones." (p.89)

Assuming a heuristic rather than algorithmic approach, descriptions of the online search process given earlier in this review (e.g., Fenichel, 1980; Bates, 1979) can be viewed within the framework of Newell and Simon's five-step process. For online searching, step 1 could involve analyzing the query in terms of the concepts involved, selecting the search system and choosing the databases to be searched. Step 2 could involve selecting an overall search strategy (for instance, a pearl-growing approach, where the searcher tries to retrieve one or more highly relevant citations and uses the information gained regarding authors, free-text terms and indexing to find other citations), and choosing appropriate terms (free text or controlled) and classification codes. The choice of initial strategy may be a function of the query or a function of the past experience of the searcher, or a combination of both. In step 3, the method could be applied by typing commands, reviewing sets, and selecting

specific search tactics to apply, such as truncation and addition or deletion of synonyms. At step 4, the searcher would decide whether to continue with a selected strategy, choose another strategy or discontinue the search. Finally, step 5 would apply in solving problems that might arise in carrying out a strategy, such as determining the best way to approach a concept that is retrieving far too many references. The moves made in the execution of the third, fourth and fifth steps of the Newell and Simon process would identify the level of heuristic approach exhibited by the searcher. A low heuristic approach would tend to result in a minimization of the last three steps, where the searcher would be interacting with the system and determining subsequent moves based upon system feedback.

Morris et al. (1989) describe an attempt to develop a prototype expert system for modifying online search strategies. They point out that "because the performance of the database is never totally predictable the search is unlikely to go exactly as planned. Most textbooks agree that good searching is heuristic." (p.415)

It appears that online searching can be made to fit Newell and Simon's theory quite well. Although it is not the intention of this study, it would be useful to test this five-step theory on the task of online retrieval to determine whether searchers apply the steps in the manner and order proposed.

Decision Making

Decision making is an integral part of problem solving, involving selecting among a set of alternatives at different points in the process to arrive at the correct solution. The study of decision making tends to be either normative, specifying what the problem solver should do, or descriptive, identifying how people arrive at decisions (Reed, 1982).

One technique that has been used successfully in studying decision making is that of means-end analysis. This type of analysis was used in the development of the General Problem Solver - a program for solving transformation type problems (problems for which the goal state is specified). The solver must generate a solution by selecting appropriate operators within the problem space that will result in successive transformations from the initial state to the goal state. A common way of representing the problem solution is through use of solution or decision trees, as described by Groner et al. (1983):

> In a first stage of problem solving, the subject has to find out what is given, what the problem is, and which actions, transformations, or operators can be applied. These operators are capable of changing the states of the problem. A problem state is defined as the initial problem situation or any state that can be achieved through the successive application of allowed operators. The set of all possible sequences of operators can be represented by a solution tree, where the branches

> represent the application of an operator and the nodes represent problem states. A solution is defined by certain properties, and all states of the solution tree that have these properties are solution states. Finding a solution then corresponds to a search for a solution state in the solution tree. (p.13)

Fidel (1986) successfully developed a decision tree of the steps involved in selecting search terms. She found that experienced searchers used a relatively well-defined set of rules for deciding whether a free text or thesaurus term would be best, based for instance, on perceived commonality of the term in the database and specialized meanings for particular concepts.

Decisions made under conditions of uncertainty pose considerable difficulty in that they require that the problem solver make estimates of the probability of events occurring (Reed, 1982). As mentioned earlier, Blackshaw and Fischhoff (1988) found that users of online public catalogues tended to be over confident for all but the easiest of tasks in their estimated likelihood of succeeding at a search. It would be useful to repeat this study with online searchers having differing types and levels of experience to determine whether experience correlates with ability to make more accurate likelihood estimates.

Given this model, which describes the stages in the problem solution at which decisions are made, it is useful to examine the types of decisions that are possible at each

stage. Newell and Simon postulate four principal kinds of decision (p.826):

1. At a knowledge state (a node in the problem space), to select an operator to be applied.
2. At a new knowledge state, to determine whether problem solving shall continue from this state or not.
3. At a knowledge state, to determine whether the knowledge state shall be remembered, so that return can be made to it at some later time.
4. At the decision to abandon a knowledge state, instead of continuing to search from it, to select another knowledge state as the backup state.

In online searching, a knowledge state could be viewed as a retrieved set, representing either the primary goal or a sub-goal. The searcher must apply an operator to change that knowledge state. In searching there is a well-defined set of commands used in conjunction with a virtually unlimited set of terms that form the set of operators available in the problem space to change the knowledge state. A major decision on the part of the searcher is to choose the appropriate operator to further a search by forming a new knowledge state. Deciding whether to remember a particular knowledge state (or set) and the sequence of operators that led to it presents sometimes excessive demands on short term memory capacity. The searcher feels even greater stress if the search is

conducted under cost constraints, in an unfamiliar subject area, or on an unfamiliar database, for example. Each decision at the terminal takes time and costs money. The decision to abandon a knowledge state, or answer set, and the strategy that has produced it also applies to online searching. Recalling Blair's (1980) research on searching biases, a good deal of cognitive strain is incurred in abandoning "anchor" sets created with an initial strategy, such sets generally representing the searcher's best guess as to appropriate operators for a particular request. Abandoning sets, and therefore strategies, must result in cognitive strain and uncertainty for the searcher.

Kuhlthau et al. (1988) used Personal Construct Theory as a framework for studying the search process. The theory describes learning as a series of phases, involving both intellect and emotions, for construct building. "In the first encounter with a new experience or idea, typically individuals are confused and anxious. This state of uncertainty builds until a threshold of choice is reached where the quest for meaning is either abandoned or a hypothesis is formed, moving the process along to confirm or reject the new concept" (p.70).

The decisions made in the course of searching, then, may follow those described by Newell and Simon and suggest points of cognitive strain in the search process. One way to reduce the strain may be to avoid assessing the knowledge state at all (viewing sets), which could serve as

an explanation of actual behaviour described for online searchers. Cognitive strain may also be reduced in the manner suggested by Blair (1980) by adhering to an "anchor set." Finally, cognitive strain is introduced when searchers must make estimates of the likelihood of events occurring, such as the likelihood of specific indexing terms being used. It has been demonstrated that, on complex tasks, searchers tend to be overconfident in their estimates of such occurrences. When that happens the searcher is faced with an unanticipated decision, such as being confronted with no retrieved items for a term that the person was confident would occur in the database. All of these aspects of decision making as they relate to online searching merit further study and may explain some of the counter-intuitive search behaviour observed.

Solving Complex Problems

Some of the key work in the area of complex problem solving behaviour was done by Herbert Simon in the 1950's. He investigated problem solving in terms of decision making and organizational behaviour. Two concepts of Simon's are of particular interest when studying interactions with complex systems.

Simon (1957) described a principle of bounded rationality, which "takes into account empirical limits on human rationality, of its finiteness in comparison

with the complexities of the world with which it must cope" (p.198). When translating experimental behaviour and problem solving theory to the real world, it is important to keep in mind this principle. "The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world - or even for a reasonable approximation to such objective rationality" (p.198). In consequence, the problem solver must construct a simplified model of the real situation in order to deal with it. His behaviour may be rational with respect to the model, but may not be even approximately optimal with respect to the real world. "To predict his behavior, we must understand the way in which this simplified model is constructed, and its construction will certainly be related to his psychological properties as a perceiving, thinking, and learning animal" (p.199).

In addition to this principle of bounded rationality, a second factor may govern, in part, the behaviour of a problem solver. This is the tendency to satisfice as opposed to optimize, in formulating alternatives to solving a problem. March and Simon (1958) explain that "most human decision-making, whether individual or organizational, is concerned with the discovery and selection of satisfactory

alternatives; only in exceptional cases is it concerned with the discovery and selection of optimal alternatives... To optimize requires processes several orders of magnitude more complex than those required to satisfice. An example is the difference between searching a haystack to find the *sharpest* needle in it and searching the haystack to find a needle sharp enough to sew with" (p.140-141).

Thus, it would be expected that subjects placed in a problem solving situation with a complex system will exhibit deficiencies in problem solving that relate to formulating and responding to a simplified, and possibly defective, model of the system. They should also be expected to seek alternatives that will lead to a solution meeting some minimum set of criteria, as opposed to optimal criteria, which should be kept in mind in studies of online searching that set retrieval goals for the searchers. According to these principles they will seek to find a minimum satisfactory response (acceptable recall and precision) as opposed to having a goal of 100% recall and precision.

Dorner (1983) describes heuristics as "the science of finding solutions to problems whenever there are no algorithms... heurisms do not offer an a priori guarantee that a solution can be found, if one exists. Heurisms increase the probability that a solution will be discovered, although the degree to which this

probability is increased cannot be stated absolutely" (p.89). The study of heuristics has been based almost exclusively on well-defined problems. These are problems characterized as: (1) having an aspired goal state exactly known with respect to the criteria that must be met, (2) having a precisely known field of reality (a network of possible facts), (3) having a limitation to static objects, such that the structure of the problem changes only when the problem solver performs an action, (4) having a limitation to not very complex objects, e.g., compare the situation of describing a state in a chess game to that of describing an economic or political situation, and (5) having a limitation of dealing with completely transparent objects with fully evident and easily conceivable characteristics.

However, as Dorner points out, most actual problem solving situations in every day life do not exhibit these characteristics; these are the complex problems which are far more difficult to study and have consequently received very little empirical study.

In working with complex systems, "lack of knowledge as to the possible operators and uncertainty as to the possible states or situations of the system is the rule rather than the exception" (p.92). The problem solver is coping with "the problems of uncertainty, lack of knowledge, and partial or imperfect knowledge" (p.92). In

addition to having vaguely defined goals, the person may be contending with conflicting multiple goals, and a system state that, at any given moment, is not readily knowable (Dorner, 1983).

It would seem that online searching as a problem solving activity would fall somewhere between well-defined problems and complex problems, i.e., a moderately complex problem. Although it does not have an exactly known goal state or a precisely defined field of reality, it does have a relatively constrained set of operators from which the problem solver can select, and the problem solver has the potential of assessing the current state of the system at any time by viewing retrieved sets. The multiple goals situation is an aspect of complex problems that does apply to online searching in that there is strong evidence that attaining the best answer set conflicts with reducing the cost of the search. Conflicting goals may also arise when deciding which concepts to emphasize in searching when several concepts seem to conflict, for instance, when combining sets results in an answer set with no hits. If the searcher is aware of the inverse relationship between precision and recall, then another set of conflicting goals may exist as well.

Dorner and colleagues have investigated solving of complex problems through creation of games based on small businesses, third world economics, and other such complex systems. They predict that subjects will make goals

precise through dimensionalizing and giving a valence to unprecise goals, i.e., describe the vaguely defined goal state in terms of relatively concrete variables. This often requires an improved understanding of the field of reality by gaining structural knowledge of the system, which is achieved through information gathering by questioning and reading, forming analogies, and application of abstract structural schemata. An important part of the activity is to choose the appropriate level of resolution for understanding the system:

> Things can be examined either too superficially or too exactly. A resolution level that is too superficial implies an undifferentiated point of view and, accordingly, an inability to plan measures adequately. A resolution level that is too detailed contains the risk of dissipating one's energies and concentrating on unimportant details. The correct level of analysis results from the choice of goal. In the terminology used here, a goal is always the set point of a critical variable. (p.97)

The last general component of finding solutions in complex systems is dealing with contradictory goals. A main deficiency of human problem solving is failure to recognize that such contradictions even exist. "A fair amount of knowledge of the section of reality is a prerequisite for discerning contradictions. Balancing goals also presumes knowledge of what is at all possible in the corresponding section of reality" (p.98).

What did Dorner and his colleagues discover about goal handling in complex systems? Basically, that problem

solvers sometimes follow these rules and sometimes they don't. The researchers described six ways in which subjects deviate from the correct way to solve complex problems, as described above. They would 'redefine the goal' by neglecting to make the goals precise enough. They would 'forget the final goal' by fixing on an intermediate goal to the detriment of solving the original problem. Some subjects neglected to take up any goals at all, but merely 'muddled through', attacking problems apparently randomly, i.e., a planless action. Another deviation was 'encapsulation', in which the subject adheres resolutely to a theme, forgetting the background and expending far too much time and energy beyond the importance of the theme to the overall problem. Another behaviour exhibited was termed 'thematic vagabonding', wherein subjects changed their area of pursuit rapidly, swinging from theme to theme instead of deciding on a course of action and seeing it to completion. The final deficiency in approach was termed 'dogmatic entrenchment':

> A deficiency that is manifested when individuals have to deal with insufficient amounts of knowledge is the abandonment of attempts to increase their knowledge and the substitution of an information-gathering process by a dogmatically entrenched system of assumptions about the section of reality in question. The individual stops the collection of data or collects only those data that fit into his system of assumptions about reality. That means that the individual never again gets negative feedback; his system of assumptions becomes dogmatic. (p.102)

Such deficiencies in goal-handling can be readily described for the task of online searching, as well. For instance, novice searchers may well have a tendency to 'muddle through' in a relatively planless approach, hoping that the course of interacting with the system will produce a solution to their problem. 'Thematic vagabonding' would also seem an apt description for what Fenichel (1980) described as switching from searching one concept to another without completing a search for a particular concept - what she describes as a departure from the postulated problem reduction approach to searching.

Dorner draws a conclusion of significance to the study of information gathering behaviour in general:

> The psychological reasons for such cognitive degenerations may often be a feeling of incompetence, a feeling of being challenged way beyond one's own ability. Indeed, the gathering of information presumes that one admits at least to oneself that one does not know enough, that the information is not available that is needed before taking action. Admitting this fact presumes a feeling of competence that is to a certain extent intact. If this is not the case, a feeling of competence can be secured by dogmatizing one's own views of the world and forgoing the gathering of information or by limiting the input of information to those data that are in accordance with one's own hypotheses. (p.102)

It has been shown that online searching bears several attributes of what Dorner refers to as a complex problem, in that it lacks a known goal state or precisely defined field of reality, and requires the problem solver to deal with multiple and possibly conflicting goals. As such, his

description of deficiencies observed in solving complex problems may show promise for the analysis of the behaviour of online searchers. In particular, it may help to explain the non-interactivity observed in various studies of online searching as a reduction of information gathering behaviour when faced with a difficult search - an unwillingness to collect information that may not be in accordance with the searcher's hypotheses about the problem state.

Individual Differences

The theories of human problem solving outlined above assume two levels of commonality in organization among all problem solvers: (1) all humans are information processing systems and therefore have certain basic organizational features in common, and (2) all humans have a few universal structural characteristics, such as nearly identical memory constraints, that tend to produce commonalities of behaviour in problem solving. Beyond this, however, there are basic differences, such as in what is stored in long term memory (e.g., knowledge about the task of online searching and about the search query), in experience and in culture.

Streufert and Streufert (1978) developed a general theory of human problem solving that may help to explain some of the variability reported in problem solving tasks such as online searching. Their theory of the response of

organisms to incongruity in the environment is particularly useful.

Individuals develop expectations about the amount of incongruity expected in specific areas or situations based on past experience. These situational expectations become generalized and abstracted into a general incongruity adaptation level (GIAL) that is specific to each individual. "Organisms with past experience rich in general incongruity would develop high GIALs, and those with relatively constant pasts would evolve low GIALs. Both would define points or ranges where the incongruity to which they may be exposed would be experienced as consistent" (p.173). This includes too little incongruity as well as too much incongruity.

The theory is of particular interest and relevance in its application to information search behaviour. Theorists have developed various models attempting to relate degree of information search to an optimal incongruity concept. The basic theory is that information search increases in both directions from the optimal incongruity level, i.e., too little and too much incongruity will prompt information search. However, past a certain point in either direction, information search will decline.

If the GIAL theory is true, we would expect different incongruity adaptation levels for the specific task of online searching. For example, searchers working in an unfamiliar subject area may have a somewhat lower tolerance

for incongruity than those experienced with the subject, and therefore, would exhibit generally lower information search behaviour in the course of searching than would searchers more familiar with the topic. It would be useful to identify factors that may affect information search behaviour to help explain the variability often reported for searchers, even those with similar training and experience.

Summary

It seems that problem solving research, and, in particular, the work of Streufert and Streufert on cognitive style and that of Dorner on complex problems, has much to offer to the study of online search behaviour. It appears that some of the often counter-intuitive and unexpected observations of behaviour of online searchers may find explanation through applications of these theories to the task of online searching. Several research questions that emerge from this review will be addressed in the following section.

RESEARCH QUESTIONS

At this point, it is necessary to resolve the broader issues of problem solving theory and cognitive complexity theory, as they relate to online searching performance,

into distinct research questions. Those factors that, in theory, should have the greatest impact on search behaviour must be incorporated into testable hypotheses.

A model of ideal behaviour in solving complex problems has been described that supposes that the problem solver will seek information about the state of the problem from various sources that can be applied to modify and improve the searcher's knowledge state. A further model, taking into account individual differences, states that for a given person and a given task, receptiveness to new information will be dependent upon the occurrence of optimal conditions for that task. Under less than optimal conditions, information seeking will fall to a minimum.

In a means-ends problem solving approach, the person has a clear idea of the final goal (often the case with simple problems such as puzzles), and will attempt the attainment of the goal by comparing the current state with the goal state and making moves that will reduce the distance between the two states. However, in solving more complex problems, optimal behaviour involves continually gathering information from various sources about different aspects of the problem. The appropriate goals must be identified at each step and the information gathered must be applied to achieve those goals. Conflicting goals must be recognized, and the problem solver must know when the final goal has been achieved and the problem solved.

Online searching is hampered by not having a clear final goal and has many other attributes in common with more complex problems. It is of interest, then, to learn more about what information online searchers are seeking and heeding in the course of a search, and what factors primarily affect their problem solving behaviour.

Several researchers have noted that it is not uncommon for searchers to use their initial strategy with no subsequent modifications and often with no final review of the answer set obtained (Fenichel, 1981; Wanger et al., 1976; Harter, 1984). There may be several explanations for this non-interactive approach, the most popular of which is that searchers are heeding a cost constraint and saving money by minimizing their interaction with the system.

Harter has referred to level of interactivity with the system as 'heuristic approach.' It would be erroneous to say that problem solvers may exhibit a non-heuristic approach; it is probably more correct to say that there are different levels of heuristic approach. A high heuristic approach for example, would be exemplified by a relatively greater amount of interaction between searcher and system, as the searcher uses the system to provide information about the problem state. It, therefore, involves gathering varied information about the problem, and seeking and utilizing feedback from the search system that can be applied to modify the search approach. For instance, it would involve gathering information about the

form and content of the database, information that would help to disambiguate the search query, information about the indexing used in relation to the problem subject area, and if possible, information from the person requesting the search.

Two general questions emerge from this. The first is whether factors can be identified that stimulate different levels of heuristic approach. The second question is whether, in fact, there is a correlation between heuristic level and search success.

What factors might contribute to a low heuristic search approach? The factor most often considered in the literature to cause low heuristic searching is cost consciousness. Cost consciousness may result from previous training, budgetary constraints, and policies that charge search costs back to the requestor. Whatever the underlying motivator for cost consciousness, it would be expected that high cost consciousness should lead the searcher to minimize the duration of the online interaction. Cost consciousness has been put forward to explain the relatively low levels of interaction and online modification of searches by even expert searchers. This, then, leads to the first research question.

Research Question 1: How does cost consciousness affect search performance?

If cost consciousness significantly influences search behaviour, it would be expected that this would be manifested in use of strategies that would tend to reduce time at the terminal. One such strategy might be to concentrate search effort at the pre-terminal stage, taking a low heuristic approach to searching at the terminal. With such an approach the searcher would be likely to maximize efforts to discover all terms appropriate to the concepts in the query before going online, to carry out the search at the terminal in a way that permits entering the terms as quickly as possible, to minimize any online modifications to the search and, possibly, to attempt to keep the answer set to a smaller size to minimize printing costs.

If search behaviour could be altered to fit the proposed model of cost conscious searching through experimental manipulation of level of cost awareness, then there would be support for explanations of non-interactive search behaviour resulting from cost considerations. In this case, cost becomes a significant factor in the search process, possibly altering the natural problem solving process.

Heuristic search behaviour could be defined in terms of high and low heuristic approach. The high heuristic approach would be characterized by being more highly interactive, and by gathering more information to apply to the problem solution. The low heuristic approach would be

characterized by being less interactive, and by gathering less information to apply to the problem solution.

This leads to the first research hypothesis. It is proposed that higher cost consciousness will result in a lower heuristic approach.

H1: High cost consciousness will result in a low heuristic search approach; low cost consciousness will result in a high heuristic search approach.

The next logical step is to consider the effect of heuristic approach on search outcome. All of the factors described above are of interest in their effect on search behaviour, and on search outcome. It is, therefore, natural to look at the relationship between heuristic approach and search outcome.

Research Question 2: How is heuristic approach related to recall?

The measure of search outcome of particular interest in this research is recall, because of the difficulty of achieving this goal. It is expected that a higher heuristic approach will result in greater search success in the form of higher recall. Thus, the following hypothesis emerges for testing:

H2: High heuristic search approach will result in higher recall; low heuristic search approach will result in lower recall.

Another factor that would be expected to influence the searcher's behaviour and ultimate success is familiarity with the search topic. This leads to the third research question.

Research Question 3: How does search topic familiarity affect search performance?

It is generally assumed that greater familiarity with a search topic will lead to a more successful search, through having a better understanding of the correct terminology, and being better able to recognize relevant documents. A searcher who is highly familiar with the subject of the query might exhibit a low heuristic approach more than would someone who needs to explore a topic and gather information while online to clarify a topic area or database coverage of a topic. Conversely, a searcher who is highly unfamiliar with a topic may take a low heuristic approach to a search because of a feeling of inability to cope with the unfamiliar 'terrain'. This would suggest that an intermediate level of familiarity with a search topic would lead to more information gathering and a high heuristic, interactive approach, as the searcher learns

about the topic in the course of the search. It is proposed that extreme levels of familiarity will reduce the heuristic approach, whereas a moderate level of familiarity will increase the heuristic approach.

H3: Extreme levels of topic familiarity (very high and very low) will result in a low heuristic search approach; moderate levels of topic familiarity will result in a high heuristic search approach.

Another factor generally assumed to influence the performance and outcome of a search is the searcher's familiarity with the database on which the search is performed. The searcher with a greater familiarity and past experience with a database will be more conversant with the indexing used and the content of that database.

Research Question 4: How does database familiarity affect search performance?

It would seem likely that all other things being equal, high familiarity with a database would lead to a less interactive approach as there would be fewer 'surprises' at the terminal. Low familiarity might cause exploratory behaviour, or on the other hand, not knowing what results to expect, the searcher might minimize the

interaction. This would seem rather subject to individual differences in approach to an unknown area. Again, it is proposed that it is in extreme versus intermediate responses on this measure that differential effects will emerge.

H4: Extreme levels of database familiarity (very high and very low) will result in a low heuristic search approach; moderate levels of database familiarity will result in a high heuristic search approach.

An important factor in any problem solving task should be level of anxiety. Anxiety adds a stress component to the problem solving process, which should affect performance.

Research Question 5: How does level of anxiety affect search performance?

Very high anxiety could depress the problem solver's ability to seek out and use new information. Moderate level of anxiety could allow the problem solver to cope with more information and still maintain a higher level of motivation to accomplish the goal. Very low anxiety might lead to low motivation about the task. This, then, is another potential factor influencing the search process.

Anxiety is a factor that should correlate with the other factors proposed to affect search performance in this study. Stresses on any of the other factors: cost consciousness, familiarity with search topic, and familiarity with the database would presumably increase anxiety.

H5: Extreme levels of anxiety (very high and very low) will result in low heuristic search approach; moderate levels of anxiety will result in high heuristic search approach.

SUMMARY OF HYPOTHESES

H1: High cost consciousness will result in a low heuristic search approach; low cost consciousness will result in a high heuristic search approach.

H2: Hign heuristic search approach will result in higher recall; low heuristic search approach will result in lower recall.

H3: Extreme levels of topic familiarity (very high and very low) will result in a low heuristic search approach; moderate levels of topic familiarity will result in a high heuristic search approach.

H4: Extreme levels of database familiarity (very high and very low) will result in a low heuristic search approach; moderate levels of database familiarity will result in a high heuristic search approach.

H5: Extreme levels of anxiety (very high and very low) will result in low heuristic search approach; moderate levels of anxiety will result in high heuristic search approach.

FURTHER INVESTIGATION

For consistency with other research into online searching, traditional search process variables should also be explored with respect to relation to heuristic approach and to search success. Correlations with measures such as number of search cycles, number of commands issued, number of citations viewed, and number of Boolean operators would be useful for providing a comparison with the results of other studies of online search processes.

Certain other variables could be measured that may affect search performance, and their influence on search success and heuristic approach assessed. One that has not received much analysis in past research but that has a potentially large impact on search performance is the effect of not being able to interact with the person requesting the search, which is a very important source of

problem solving information. The impact of this on the search process was explored in this investigation.

METHODOLOGY

DESCRIPTION OF THE STUDY

The aim of this study was to look at the effect of cost consciousness, experience level, heuristic level and several other variables on search performance. An experimental design was selected in order to take a rigorous approach to measuring and manipulating variables to look at their effect on search performance. In particular, it was important to manipulate the cost consciousness variable and observe its effect on search process and outcome. The experimental approach increases internal validity, and thus provides for greater confidence that a change in an independent variable results in a change in a dependent variable. However, the tradeoff is a decrease in external validity, that is, the extent to which the findings can be generalized to other populations and other settings, in this case, to the "real world" behaviour of online searchers.

Two dichotomous independent variables were expected to affect search process and outcome: experience level (novice or expert searcher) and cost constraint (present or absent), giving rise to a two-by-two factorial design. The four resulting cells required 40 subjects, 20 experienced searchers and 20 novices. Within these two

groups, half had a cost constraint imposed and half had no such constraint. This resulted, then, in 10 subjects per cell.

Each subject performed the same three search tasks, giving rise to a total of 120 searches. The first search was used to orient the subject to the experimental situation. The second and third searches, both categorized as being of intermediate difficulty, were alternated such that 50% of the subjects searched question 2 first and 50% searched question 3 first. The experiment was conducted on the ONTAP ERIC database of the DIALOG search service.

Subjects were required to fill out a background questionnaire focussing on their education and the nature of their experience witn computers, online searching, and library activities. They also filled out brief questionnaires before and after each search, eliciting information on their feelings about the search.

Subjects conducted their searches while verbalizing their thoughts, which were recorded. A transaction log was stored for each search.

In addition, a measure called heuristic index was developed as an indicator of the heuristic approach of the searcher. This measure was then used to look at the relationship between heuristic approach and search performance and search success.

SUBJECTS

Subjects were selected to represent a range of type and amount of online experience. They were drawn from a novice pool of library and information science students, searchers in special and academic libraries, and a few expert instructors in online searching. All experts were practicing librarians and all novices were students in an M.L.S. program at one of two schools.

The novice searchers were solicited through notices placed at two library schools. They were required to have had at least one course in online searching, and to be capable of formulating and executing a search strategy on DIALOG. Novices were paid $10 to participate in the study.

The experienced subjects were solicited through direct communication with the heads of various libraries known to engage in fairly extensive online searching. A total of eleven different libraries took part in the study. The person administering the online search department selected the most appropriate subjects and approached them about participating in the research. Participation was entirely voluntary.

Willing candidates were given a background questionnaire to fill out, which was accompanied by a letter describing the research and a consent form (Appendix 1). The tasks to be performed were outlined in the letter, and the subjects were assured anonymity and the opportunity

to withdraw from the experiment at any time. Several potential subjects refused to participate because of reservations about verbalizing their thoughts and being audiotaped.

It was necessary to test 22 novices, in order to find 20 novice subjects who were able to complete all three search tasks. Two novices dropped out of the study because they were unable to finish one of the search tasks. In the expert group, 22 searchers were tested in order to find a total of 20 expert subjects. Two expert subjects had to be dropped from the study. One was dropped because the transaction log was not saved properly and further analysis of the data was impossible. The second was dropped because approximately 90% of the audiotape for the verbal protocol was inaudible.

Subjects were solicited until the four cells in the experimental design were full (10 subjects per cell). Because the subjects were located in three different geographic locations, a research assistant administered some of the search tasks. This introduces a potential source of bias to the experiment; however, the assistant was given detailed instructions on the procedures to be followed, and notes were made on any discrepancies that occurred.

COST CONSTRAINT TREATMENT

Within each of the novice and expert subject groups, half of the subjects were randomly assigned to either treatment or control groups. For the expert group, this was done by library, so that at each library approximately half of the participants were assigned to each of the treatment and control groups. This was done to randomize any effects on search performance caused by search environment. Given that different types of libraries have quite different charging policies, for example, academic libraries charge back the cost of the search to the client whereas special libraries do not, it was important to control for library type in assigning subjects to cost consciousness treatment and control groups.

The treatment took the form of an explicit instruction at the beginning of the search session that search costs would be monitored and that it was very important that cost be kept to a minimum while doing the best search they could (see Instructions to Participant, treatment and control versions, Appendix 2). Treatment subjects were also 'reminded' that they should be cost-conscious by being told to request 'costs' from DIALOG at the end of each search. If they had been given a time limit, it would be difficult to conclude that search behaviour was, in fact, altered, when it could be that they simply ran out of time before making significant modifications to the initial strategy.

The control group subjects were simply told at the outset that they should not be concerned about the cost of the search.

SEARCH TASKS

Each subject was required to conduct three searches. The same three queries were used for all subjects; however, the question order was varied. The first query was the same for all subjects. This was intended as a familiarization exercise. Within each of the four cells, the order of the second and third queries was reversed for half the searchers (five searchers). This was to reduce any bias inherent in the order of the queries and to distribute variations in search process that are due to a practice effect equally over the two search queries. For expert searchers, who were from a variety of libraries, within each work setting subjects were randomly divided between the two search question orders.

The first search question, the practice question, was taken from the 'beginner' category, and the second and third questions were taken from the 'intermediate' category of ONTAP ERIC. The intermediate level of difficulty was selected because, as has been pointed out by other researchers, significant differences in searching may be masked by search tasks that are too simple to be cognitively challenging. It was important to ensure that

subjects perceived the task as non-trivial. It was also important that the novice searchers not be overwhelmed by the search task, thus the 'advanced' category of search queries was not used. The search queries were as follows,

Search Question 1: Parapsychology.
Search Question 2: Library services for the physically handicapped (not mentally or language handicapped).
Search Question 3: White flight to the suburbs.

The tasks for analysis were questions 2 and 3. The topics were selected for two reasons. One reason for the selection was to provide topics that differed in general familiarity. The library science question would be familiar to most of the subjects, whereas the white flight question would be unfamiliar to most of them, as it is primarily an American phenomenon and the searchers were Canadian. It is also a somewhat dated topic in that it is not often heard of today. In other words, the majority of subjects may never have heard the term 'white flight' before. However, this poses an interesting challenge to the searchers; if they are to understand the topic, that understanding must come from interacting with the database. Thus, for question 3 in particular, a heuristic approach should improve search success by leading to better understanding of the search topic. The other reason for the choice of these particular topics was that question 2

provided several facets and fairly explicit terminology. Search question 3, on the other hand, employed the briefest of descriptions and used jargon terms. Question 2, on library services for the physically handicapped, had an answer set of nine citations, as defined by ONTAP ERIC. Question 3 had only five citations in the answer set. It was expected that of the two questions, question 3 would be the more difficult.

All subjects were given as a goal to find the best answer set for the search question. They were told tc "find everything on the topic;" in other words, high recall was the goal of the search. No mention was made of a precision goal; this was left up to the individual to decide what sacrifice in precision should be made to obtain high recall. The high recall goal was chosen because various performance studies indicate that it is the most difficult goal to attain, because it requires that the searcher know what is available in the database and when everything relevant has been found.

SEARCH ENVIRONMENT

ONTAP ERIC is a special training subset of the ERIC database, consisting of 35,000 references entered into the database in 1975. The search questions were developed from actual queries to the ERIC database. It has several features that made it desirable as an experimental tool.

Because answer sets are provided online, recall and precision can be calculated. In addition, it is a relatively cheap database to search, its subject coverage tends to be non-technical and does not require a great deal of specialized background knowledge, and there is a thesaurus to aid in strategy development.

The online answer sets were created by exhaustive searching of the file and represent 100% recall and 100% precision. It was decided that the ONTAP ERIC answer sets would be considered 100% complete, despite the fact that other researchers have questioned their completeness (Fenichel, 1981; Jackson, 1981). If the answer sets are lacking one or two references that seem relevant, at least all subjects will have the same handicap in recall and precision calculations. Jackson (1981) warns that searchers using the recall and precision calculation facility on ONTAP ERIC to compare retrieved sets with stored answer sets may be confused by discovering retrieved citations that are considered relevant but are treated as false drops by the system. In the current research, the subjects do not compare their retrieved sets with stored sets, and indeed, may not be aware that such sets exist. Therefore, Jackson's concern does not apply to the experimental situation. However, his concern for the completeness, or rather incompleteness, of the answer sets merits consideration. Given the difficulties involved in determining relevance, the suspected incompleteness of the

answer sets seemed a relatively minor drawback. The intention was to control environmental variables as much as possible by using the same questions, database, search system, and command language, for all subjects.

EXPERIMENTAL PROCEDURES

The procedures are described in the Instructions to Participant (Appendix 2). After welcoming the subject, the experimenter gave a brief description of the series of tasks the subject was to perform. Any problems the subject had had filling out the background questionnaire were discussed. The subject was permitted an opportunity to ask questions about the experimental tasks before beginning, but any questions about the research itself were deferred until the end of the session.

Subjects were given the DIALOG Quick Guide, the ERIC thesaurus corresponding to the year of the database (1975), and a worksheet with the search topic typed at the top. They were asked to make any notes about strategy on the worksheet, but were not required to prepare a formal strategy for submission to the experimenter. They were given as much time as they liked to prepare their strategy, and they indicated to the experimenter when they were ready to begin. The experimenter did the login and logout on DIALOG and set up the downloading procedure for capturing the transaction log.

Before beginning the first search, a talk-aloud practice exercise was administered to familiarize the subject with what was required in terms of verbalizing thoughts during the search (Appendix 2). Many subjects expressed reservations about talking aloud and being audiotaped during the search. All seemed to overcome their 'shyness' early into the first, practice, search, and many stated at the end of the session that they were surprised how little the taping had bothered them 'once they were used to it.'

The subjects were given a pre-search questionnaire to fill out before each search and a post-search questionnaire immediately following each search (Appendix 1). The questionnaires were designed to be the source of data for the hypotheses regarding anxiety level and familiarity with the search topic. They also permitted collection of data about other feelings and impressions about the search task. Several of the questions were devised to permit self-rating on a five-point Likert-type scale. An initial set of questionnaires was prepared and tested in the pilot test.

The result of the pilot test was that subjects were favouring the middle ratings, so it was decided to expand the scale to seven points to give a wider range between the extreme points and allow for finer distinctions to be made. Two questions were included about the subject's feelings regarding not being able to interact with the person requesting the search. This was done because it was

necessary to get a judgement before and after the online session, when the subject's opinion may change about the need for interaction depending upon the success of the search. The questions also permitted the subject to express frustration over the artificiality of the search task in which they were given only a very brief written description of the problem. A question was included in the post-search questionnaire about the subject's level of cost consciousness. This was included to verify the treatment effect.

Search results were recorded on a Search Data Form (Appendix 1), which was designed to allow quick recording of results and comments during the session. Search outcome measures were recorded on this form, including the number of the final set that the subject submitted as his or her answer. Any comments about irregularities in the experimental situation were recorded on this form, as well. The cost of the search was recorded on this form, which, in the case of the treatment group, served as a reinforcement of the treatment in that subjects were asked for their final costs, which were 'obviously' being recorded as part of the search data. Search cost was recorded for controls as well, but they were not aware of it.

Discrepancies in Procedures

It was apparent from the audiotapes of the sessions with the novices that there was more interaction between experimenter and subject than occurred with the expert group. The greatest level of interaction occurred with the first, familiarization search question, which was not used in the analysis anyway. Not surprisingly, many of the novices needed considerable help initially in recalling DIALOG commands. Had they not been provided this assistance, many would not have succeeded in obtaining any sort of answer set; however, it did introduce a bias in favour of the novices. It should be noted, however, that a few of the experts also needed help recalling basic DIALOG command syntax when carrying out the first search task. Observation of novice performance, without assistance, would be interesting, but was not the goal of this research. Given the general level of struggle involved in mastering the command language by the novices, the few instances of help provided probably did not significantly bias the results.

STRUCTURE OF DATA ANALYSIS

Coding the Search Transcripts

The verbal protocols were transcribed by the experimenter from the audiotapes. Transcription of audiotapes is a very time consuming process. The taping took place at the terminal in the computer lab (in the case of the novice searchers) or at the online workstation in the library setting. There was usually a large amount of background noise from the environment, including noise generated by the computer itself. Thus, the transcription often involved listening to portions of the audiotapes several times to be certain that the interpretation was correct. There tended to be either long periods of silence while subjects were reading displayed text, or very rapid verbalizations that required going over the tape several times to capture them accurately. After the initial transcription of the audiotape, the tape had to be played again while consulting the transaction log together with the transcript of the session so that search commands could be inserted into the transcript at the point where the searcher was typing in the command. This was relatively easy to do as the typing sounds were distinctly audible and the previous or subsequent utterances of the subject indicated what the subject was typing. The result was a complete log of commands and verbalizations for each

subject for the second and third search questions (80 logs in all). The first search question was not transcribed because it was a familiarization task and was not analyzed further. An example of one such complete log is given in Appendix 3.

The complete logs were then "shuffled" with novice and expert logs combined in random order and renumbered. In this way, the identities of the subjects were hidden from the coders.

The further analysis of the logs primarily involved coding of the search commands, using the verbal protocols to disambiguate the subject's intentions where necessary. The first step was to go through all of the complete logs and code commands that were errors: (1) a "logical error" was coded when the subject indicated that an error had been made and immediately corrected the error with a subsequent command, e.g., used an incorrect Boolean operator or used a wrong set number; (2) a "command error" was coded when the subject received a system error message in response to a command that the system was unable to execute, e.g., incorrect command name or unbalanced brackets. This coding was done by the experimenter.

Search terms were then coded as being either an "original" search term or a "new" term. Original terms were either (1) original terms appearing in the subject's pre-search strategy formulation or (2) stated by the subject to be a pre-search strategy term. The so-called

"new" terms were those terms added to the search as a result of strategy modification in the course of the search. The new terms were coded as (1) a modification to an original term (e.g., a term was truncated or an adjacency operator was added), (2) a term obtained from citations viewed, or (3) a term of unknown origin, appearing "spontaneously" in the course of the search. The verbalizations were essential for disambiguating errors and classifying search terms.

Coding term types required use of the pre-search formulation forms, the complete logs, and the transaction logs containing the typed citations (to determine whether a term was derived from a viewed citation). The coding is time consuming, given that there are three different sources of data to consult for each code assigned, and that the coders must continually move back and forth among these sources. The coding was done by having two independent coders code all of the search logs for search questions 2 and 3 (80 in all). In the case of disagreements, each coder gave the reasoning for the assigned code and they both reached an agreement on the best code. The instructions to the coders are given in Appendix 4.

Description of Variables

Independent Variables

Two dichotomous independent variables were used in the experiment. One was experimentally manipulated in the form of an imposed cost constraint treatment. The other was a control variable, which was experience level. Experienced subjects were working in a library setting and had been doing online searching for at least six months. Novice subjects were library and information science students who had completed at least one course in online searching.

Search Outcome Variables

Three search outcome variables, defined in Table 1, were measured. Recall and precision were used because they are standard, though imperfect, measures of search success. Recall is the outcome variable of particular interest here because it is the most difficult to attain. Precision and search costs (total cost and unit cost) are monitored because it is generally assumed that high recall is obtained by sacrificing higher precision and lower search costs.

TABLE 1: Search Outcome Variables

Variable	*Measure*
Recall	Proportion of all relevant references that are retrieved (%)
Precision	Proportion of retrieved references that are relevant (%)
Total cost	Total cost of the search ($)

Heuristic Approach Measure

An attempt was made to develop two separate measures for heuristic approach, a measure called heuristic index, which was derived from several search process variables, and a measure called heuristic level, which was to be assigned by independent coders. The two measures for heuristic approach would have been preferable to only one, as they would be measuring different aspects of heuristicity and would have provided for greater confidence that the phenomenon in question was being accurately measured. However, the measurement of heuristic level proved to be problematic, as will be described below.

The measure called 'heuristic index' was derived by creating an index from four search process variables that reflect amount of interaction with the search system for information gathering purposes: number of cycles (where a cycle is defined as a series of commands that begin with a search command and end with a type command), number of

commands issued, number of citations viewed, and number of new search terms added in the course of the search. These four counts were normalized (by subtracting the mean and dividing by the standard deviation) and an index measure was derived by averaging the four normalized values.

These variables were chosen because they reflect individually the amount of interactivity with the system, because they may reflect information gathering on the part of the searcher, and because they each had a wide enough range of scores that they provided for greater separation on the heuristic index measure. The measures for number of search cycles, number of citations viewed and number of commands issued may tend to be correlated, in that they each reflect interactiveness with the system. However, it might be expected that the strength of correlation would vary with the searcher and with the search question. Depending on the style of the searcher, one command may be used to view a number of citations, or a separate command may be used to view each citation. Similarly, many citations may be viewed in each cycle or one citation may be viewed in each cycle. The number of new terms added in the course of the search would not be expected to be correlated with the other measures. Again, depending on the style of the searcher, each new term added may represent a new command issued, or several new terms may be added with a single command.

It has long been assumed that a greater number of search cycles and a greater number of citations viewed in the course of the search will lead to higher recall and greater search success, but this is an assumption that has not been tested. It was the intention of this research to test assumption indirectly, by including these variables in the heuristic index measure. It has not been shown that higher values on any of these measures necessarily correlates with greater search success. For instance, searchers may view retrieved citations for a number of reasons, none of which may actually relate to improving the recall obtained.

The measure called 'heuristic level' was to be derived by having independent coders assign a rating of '1' through '7' for each subject for each search, where '1' represented 'low heuristic level' and '7' represented 'high heuristic level.' The coders were given a definition of heuristicity and heuristic level. They were also given several examples that were extracted from the pilot test search logs. The coders were to use the complete search logs and the transaction logs to assign a value for each search, for each subject. The coders were required to study the examples and the definitions until they felt confident that they could code the search. For each complete log, they had to read through the entire log, referring to the printout of the transaction log whenever a searcher typed out citations. From this they were to gain an overall

impression of the heuristic approach of the searcher. This tended to be more straightforward for shorter searches than for longer ones. For some logs, the agreement between searchers was high; these tended to be clearcut examples of either very high or very low heuristic approach. For instance, very short logs usually corresponded to very low heuristic approach. However, very long logs did not necessarily correspond to high heuristic approach as searchers may have been doing little to actually gain useful feedback from the system. For example, they may have spent an inordinate amount of time looking at brief citations, without the associated indexing, and attempting to make only relevance judgements.

Three separate attempts were made by two coders to code a random sample of logs. The instructions were re-written and made more explicit after the first and second attempts, but there was still insufficient agreement between coders to ensure confidence in the the validity of the measure (coefficient of reliability measures were below .5). After the third coding attempt, the measure was abandoned, and it became necessary to rely exclusively on the heuristic index measure. It was unfortunate that the heuristic level measure did not work as planned; however, it is also interesting to note the difficulty of developing a quantitative measure for a concept such as heuristic approach to online searching.

It is difficult to integrate into one measure the various aspects of heuristic approach. If a subject uses one method of obtaining feedback from the system to the exclusion of all others, does that constitute a high heuristic approach? Can different types of interactions and attempts to gain feedback be weighted or do all essentially have the same value? These are some of the problems involved in developing a measure of heuristicity. It is encouraging that the coders seemed to understand the concept being measured. The problem lies in clearly defining the various components of heuristicity and then making decisions as to how to integrate them into a single measure. Given the importance of heuristic approach to effective problem solving, developing a valid measure for the phenomenon would provide a valuable contribution to research in this field.

Search Process Variables

Certain measures have been made in past research that are intended to provide information on the search process. These are relatively simple counts of activities occurring during the search. Nine such variables are included in Table 2.

Simple counts were made of number of each type of error, number of each type of command issued (minus those issued as a compensation for an error), number of sets

retrieved (minus those created as a result of an error), number of citations displayed number of sets from which citations were viewed. Various other counts were also made such as number of each type of search term (descriptor, identifier, free text) and number of Boolean operators.

Number of cycles was counted for each search. A cycle was defined as a series of commands ending in a Type or Display command. A new cycle was begun with any other command or with a Type or Display command applied to a different set number.

The search process variables were recorded as counts of each type of command entered that was not previously coded as an error. Complexity of the final search formulation was determined for each search, for each subject (80 in all). This measure was derived by starting with the formulation giving rise to the final set and working back through the transaction log, substituting search terms and operators wherever set numbers were used. The result was a string of terms, operators (Boolean, adjacency, and limits), and parentheses, which were counted to give the complexity measure for the search. Deriving this measure is a relatively time consuming process, requiring that the coder pay close attention to the transaction logs.

TABLE 2: Search Process Variables

Variable	*Measured By Counts Of*
Boolean ORs	
Boolean ANDs	
Commands issued	(SELECT + SSTEPS + COMBINE + TYPE + EXPAND + DS)
Original terms	From pre-search formulation
New search terms	Modifications to original terms + Terms found in citations + Spontaneous terms from unknown source
Citations displayed	Any format
Sets browsed	Sets from which citations were displayed
Cycles	Series of commands ending in TYPE (two TYPE commands in sequence count as a cycle only if they display citations from different sets)
Complexity	Of formulation yielding answer set (Adjacency + Parentheses + Truncation + Limits + ANDs + ORs + NOTs)

Self-rated Variables

Certain impressions held by the subjects were measured for each of the searches they performed, and are described in Table 3. These were measured by means of pre- and

post-search questionnaires. Most were made on a seven-point Likert type scale (see Appendix 1).

TABLE 3: Other Search-specific Variables

Pre-search Questions

Search topic familiarity (1=not familiar, 7=very familiar)

Lack of requestor interaction (1=not detrimental, 7=very detrimental to the search)

Some idea of number of relevant references expected (yes/no)

Using best search terms (1=not confident, 7=very confident)

Post-search Questions

Obtained all relevant citations (1=not confident, 7=very confident)

Content of answer set (1=not satisfactory, 7=very satisfactory)

Size of answer set (too large, too small, about right)

Level of anxiety during search (1=not anxious, 7=very anxious)

Cost consciousness during search (1=not concerned, 7=very concerned)

More requestor interaction would improve search (1=not at all, 7=very much)

PILOT TEST

Six subjects were studied in the pilot test, representing a range of online searching experience. The purpose of the pilot test was to gather exploratory verbal protocols from which a coding system could be developed, and to verify the methodology to be used. It was necessary to ensure that the verbal transcripts could be matched to the transaction logs, so commands could be inserted into the verbal transcripts at the correct place. This turned out to be relatively easy to do. Each pretest subject was given two searches to perform on ONTAP ERIC. These two searches correspond to the second and third search tasks for experimental subjects.

From the pilot test, it became obvious that an initial, simple search task should be given that would permit any problems with the hardware and software at the search site to be removed. It would serve to familiarize the subject with the DIALOG interface and with the unique and rather artificial aspects of the search environment. Finally, it would help the subject to overcome initial nervousness about talking aloud, being audiotaped, and having an observer in the room during the search.

The talk aloud practice questions were modified following the pilot test. The practice exercise initially consisted of a multiplication task and an anagram to be worked 'in their heads' verbalizing their thoughts while

they solved the problem. It became apparent, however, during the pilot test that subjects were unable to verbalize their thoughts while working the anagram, so it was dropped. The multiplication practice task did work quite well.

The pre- and post-search questionnaires were modified as described above. The five-point scale, where used, was changed to a seven-point scale to give more range to allow subjects to make finer distinctions.

RESULTS

INTRODUCTION

This chapter is divided into three sections. The first section describes checks made on procedures, which includes a description of the experimental subjects. The results of the coding reliability tests are given in this first section, as are the results of the tests of treatment effectiveness. The derivation of the heuristic index variable is described at the end of this section.

The second section gives the results of the tests of the research hypotheses. A brief description of the inferential statistics applied to the data is given initially. Following this, each of the hypotheses is tested in turn.

The third and last section presents results of additional data analyses. In this section, relationships among variables other than those in the research hypotheses are investigated. The relationships among the various self-rated variables and among the search process, search outcome and heuristic level variables are explored.

PROCEDURAL CHECKS

Description of Experimental Subjects

Subjects filled out a background questionnaire from which various demographic characteristics were measured. The experts and novices were compared on these questions and the results are given below.

In terms of gender, the proportion of females was somewhat greater for the expert group. Eighty-eight percent of the expert group were female, whereas 61% of the novices were female. Four of the expert group and two of the novice group chose not to divulge information on their gender.

Experts and novices were identical in terms of previous degrees, with 85% having a Bachelor's degree and 15% having a Master's degree as the highest degree before the MLS. The mean date that the highest degree was acquired was 1973 for the experts and 1983 for the novices. The breakdown by degree major for the two groups is given in Table 4.

TABLE 4: Frequency of Degree Major, by Experience Level

Degree Major	*Experts*	*Novices*
Natural science	4 (20%)	4 (20%)
Social science	7 (35%)	6 (30%)
Arts	7 (35%)	7 (35%)
Other	1 (5%)	1 (5%)
Missing	1 (5%)	2 (10%)

Experts and novices were nearly identical in response to the question of whether they are regular computers users. Eighty-five percent of experts responded 'yes' and 90% of novices responded 'yes.' Median number of computer courses taken was two for both experts and novices. The nature of the regular computer use is given in Table 5.

TABLE 5: Nature of Regular Computer Use, by Experience Level

Type of Computer Use	*Experts ('yes' response)*	*Novices*
Coursework	10%	90%
Games	5%	20%
Text-editing	50%	85%
Programming	5%	30%
Statistics	10%	10%
Online search	90%	60%
Other use	20%	20%

Subjects were asked to rate themselves on the extent of their manual search experience (1=not experienced and 7=very experienced). The result was that experts had a median of six and novices had a median of five.

Subjects were also asked to rate themselves on the extent of their online experience (1=not experienced and 7=very experienced). The result was that experts had a median rating of six (mode of six) and novices had a median rating of three (mode of four). One hundred percent of the experts had performed an online search within the last month, whereas 68% of the novices had done a search in the last month and 32% within the last six months. The mean number of searches per year for experts was 339.5 (median 240); the mean searches per year for novices was 38.0 (median 24).

Subjects were asked to rate the proportion of all their searching that is done on the DIALOG search service and on the ERIC database (<10%, 10-50%, 51-80%, >80%). Some subjects wrote in a zero category, rather than using the <10% category. For proportion of searching on DIALOG, experts had a mode of 10-50%. Novices had a mode of >80%. This indicates that of the relatively small amount of search experience that the novices had, most of it had been on DIALOG. For proportion of searching done on the ERIC database, both experts and novices had a mode of <10%.

Subjects were asked whether or not they had ever used the ERIC database and the ONTAP ERIC training database.

For the ERIC database, 80% of both experts and novices had used it at least once. For ONTAP ERIC, 55% of the experts and 100% of the novices had never used it.

The results would seem to verify that experts had a wider range and greater length of experience than did novices. Experts were generally older than novices. No attempt had been made to match subjects in terms of age. In terms of degree major, there was no great difference in subject background. The two groups were equally matched on regular computer use, although the nature of that use differed. They were also relatively well matched in terms of manual search experience. Experts had more recent online search experience and reported doing far more searches per year than the novices. Of the online experience that novices had had, relatively more of it was on DIALOG than was that of the experts, who were using different search services from day to day. Few subjects in either group did regular searching on the ERIC database. None of the novices had ever used ONTAP ERIC, whereas 45% of the experts had used it.

Coding Reliability

The origin of the search terms used in the course of the search was determined by coding the search terms according to the classification given in Appendix 4 (Coding the Search Transcripts). Coders used the complete logs,

the search question forms on which the searchers had noted their strategy before beginning the search, and the transaction logs. The transaction logs were necessary to determine whether a new term was derived from viewed citations. The coefficient of reliability for question 2 was .72, and for question 3 was .79. This was calculated by dividing the number of coding decisions that were in agreement by the total number of coding decisions made. In the case of disagreements, coders discussed the reasons for their decision and reached a joint decision.

Treatment Effectiveness

To determine whether the treatment imposed had an effect on the subjects' cost consciousness, the self-reported cost consciousness was used as a check. Given that experience level might have an effect on cost consciousness and on susceptibility to the treatment, experience level was controlled for in this test.

Subjects were asked to report their level of cost consciousness on the post-search questionnaire. They rated themselves on a seven-point scale (1=not concerned, 7=very concerned). The result was the distribution given in Table 6, in which nearly 50% of the subjects, for each search question, rated themselves as 1 (not concerned about cost).

TABLE 6: Frequency Distribution of Self-rated Cost Consciousness

Self-rated Cost Consciousness	*Value*	*Frequency Question 2*	*Frequency Question 3*
Not Concerned	1	17	14
	2	8	9
	3	2	3
	4	5	3
	5	4	5
	6	4	4
Very Concerned	7	-	1
	Missing	-	1
	Total	40	40

To check for treatment effectiveness, the self-rated cost consciousness was treated as the dependent variable and an Anova test was performed using treatment condition, experience level and question number as the independent variables. The cost consciousness treatment was the only variable having a significant effect on self-reported cost consciousness (F=17.246, P=.000). The mean for the treatment group was 3.48, and for the control group was 1.87. Therefore, it appears that the treatment was effective in eliciting a feeling of cost consciousness.

When the effect of experience level, cost consciousness treatment, and search question was tested on the dependent variable total search cost using Anova, cost consciousness treatment emerged as the only significant main effect (F=12.54, p=.003). The mean for the treatment

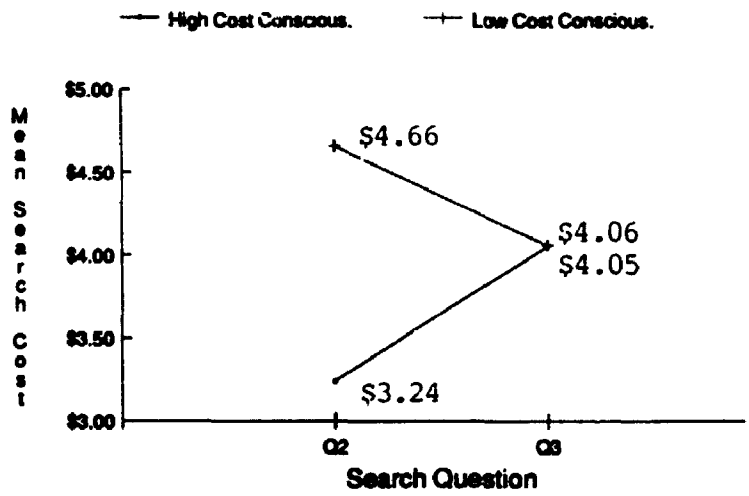
group was $2.87 (SD=1.05), and for the control group was $4.83 (SD=2.22). This provides further evidence that the cost consciousness treatment imposed on subjects in this study did have an effect on search behaviour that resulted in lower search costs. Because no significant effect was seen for experience level, it appears that the treatment was effective for both experts and novices.

The results of these tests were encouraging in that they provided evidence that the cost consciousness treatment imposed on the subjects was effective in eliciting both a heightened concern for search cost and a reduction in overall search cost. This effect was not significantly different for novice and expert searchers.

It is important to know whether subjects responded to a feeling of cost consciousness by searching in a way that resulted in lower search costs. To check this, the self-rated cost consciousness was used to divide subjects into low and high cost consciousness groups, using a median split. One group consisted of subjects rating themselves as 'not concerned about cost' (self-rated at '1'); the other group had some concern for search cost (self-rated at '2' through '7'). The resulting dichotomous variable was used in an Anova test along with experience level and search question to determine the effect on total search cost.

The result was a two-way interaction between cost consciousness and search question (F=4.65, p=.038). The means are given in Figure 1, below.

FIGURE 1: Mean Search Cost for Groups Formed from Self-rated Cost Consciousness Versus Search Question



The group rating itself as not concerned about search cost had a higher mean search cost for question 2 than for question 3. This was reversed for the group rating itself as having some concern for cost. It is interesting that mean search cost for both cost consciousness groups was virtually identical for question 3. However, for question 2, there was a much greater range in the means for the two groups. It is possible to speculate that under cost pressure, subjects were able to increase their efficiency without sacrificing effectiveness for question 2, which was

the more straight-forward question. However, for question 3, they may not have been able to increase the efficiency of their searching without sacrificing effectiveness. Another possible explanation for the lack of effect of cost consciousness on search cost for question 3 is that given their difficulty in understanding the topic, even the subjects without a cost pressure may have simply run out of ideas to try to improve the search.

Measurement of Heuristic Approach

Measurement of heuristic approach was operationalized by the development of a heuristic index measure. The heuristic index measure was derived from four traditional measures of search process (number of new terms introduced in the course of the search, number of search cycles, number of commands issued, and number of citations viewed) that were normalized for each search question (by subtracting the mean for the question and dividing by the standard deviation), and were averaged to give a single value for each subject, for each search. These four measures were chosen because they are somewhat crude measures of the amount of interactiveness with the system, as well as attempts to obtain feedback from the system to improve the search, or at least to keep it on track. One criterion for inclusion of a variable in the measure was that there be a wide enough range of scores on the variable

to give sufficient separation on the heuristic index. Descriptive statistics for this measure are given in Appendix 5, Table A1. It would be expected that there would be some correlation between cycles, citations viewed and number of commands issued, but not necessarily between number of new terms added and the other three variables. The correlations among the variables used in developing the heuristic index measure are given below in Table 7.

TABLE 7: Correlations Among Four Variables Used to Develop the Heuristic Index Measure

	Cycles	*Citations Viewed*	*Commands*
Question 2			
Cycles	-	-	-
Citations Viewed	.6498	-	-
Commands Issued	.6237	.5695	-
New Terms Added	.1948	.3257	.6216
Question 3			
Cycles	-	-	-
Citations Viewed	.4901	-	-
Commands Issued	.6460	.3991	-
New Terms Added	.3211	.1182	.8840

It does appear that there are relatively strong correlations among some of the variables used to create the heuristic index measure. However, it appears that the strengths of these correlations vary for the two search

questions. It is also clear that the number of new terms added does correlate positively with the number of commands issued; however, the correlation with the number of search cycles and number of citations viewed is not significant.

Given that the relationship between the heuristic index measure and recall is central to the second research hypothesis, it is important to look for correlations between recall and the individual measures that make up heuristic index. Certain assumptions are made about the process of online searching that suggest that recall is positively correlated with search process measures such as number of citations viewed and number of search cycles. The correlations between recall and precision and the four measures that make up the heuristic index variable are given below in Table 8.

TABLE 8: Correlations Between Heuristic Index Composite Variables and Recall and Precision

	Cycles	Cites Viewed	Total Commands	New Terms
Question 2				
Recall	.2526	.3997	.3867	.1929
Precision	.0629	.0989	.0562	-.1563
Question 3				
Recall	.0942	.4173	.0683	-.0223
Precision	.0484	.0987	-.1029	-.2165

It is clear from Table 8 that no strong correlations exist between recall and the variables comprising the heuristic index measure. There is a low to moderate positive relationship between recall and number of citations viewed for both search questions, and a low positive correlation between recall and number of commands issued for question 2. Therefore, it appears that recall, in fact, is not strongly correlated with these search process measures. The relationship between recall and heuristic index will be described in the following section on testing the research hypotheses.

HYPOTHESIS TESTING

Statistical Approach

Descriptive statistics for the search outcome, search process, self-rated and heuristic variables are given in Appendix 5, Tables A1 through A4. Because recall and precision were non-normally distributed and violated the assumptions of the Anova test, they were transformed using a square root arcsine transformation to stabilize the variances. This transformation is used when the data are in the form of a proportion, percentage or rate. The resulting transformed variables are included in the descriptive statistics in Appendix 5, Table A1.

The primary focus of this research was to identify factors affecting recall, to which end the experimental task was linked to a goal of high recall. However, it cannot be ignored that there are other factors influencing search success. In particular, a goal of attaining some level of precision is inherent in the search task; otherwise, one could simply retrieve the entire database in response to a request. In addition to this, online searching is a costly process and most searchers in "real world" settings must be cognizant of this fact. Because of this, where recall served as the dependent variable, tests of the same independent variables on precision and cost are included as well.

The Anova test was used where possible because it allows for the incorporation of several variables into a single analysis of variance, which reduces the overall alpha error probability for the test. In all analyses, the .05 level of significance was the criterion chosen.

Subjects were not given any indication as to the difficulty of the search tasks, as defined by ONTAP ERIC. As mentioned earlier (in the Methodology chapter), it was expected that subjects would find question 3 the more difficult question to search, although both questions were from the intermediate level category. The search question was included as an independent variable in the hypothesis tests, and was treated as a within subjects factor in the Anova design.

Hypothesis Tests

Test of H1: High cost consciousness will result in a low heuristic search approach; low cost consciousness will result in a high heuristic search approach.

This hypothesis was tested statistically with Anova using the MANOVA procedure in the SPSSx statistical package (Statistical Package for the Social Sciences, Version X). A design was used which had subject nested within cost consciousness treatment and experience level, and crossed with search question. This approach allows the effects of treatment and experience to be assessed independently of the question, and indicates interactions among all the independent variables, including any interactions with the question variable. The dependent variable was heuristic index.

The Anova test with heuristic index as the dependent variable resulted in no significant main effects or interactions. Therefore, hypothesis H1 that high cost consciousness will result in a lower heuristic approach is not supported.

The frequency distribution for the variable self-rated cost consciousness gave rise to a concern that a few subjects were not responding to the treatment and control conditions, and that this might obscure the true effects of cost consciousness on the dependent variables. Therefore,

it seemed prudent to verify the Anova tests described above using self-rated cost consciousness in place of the treatment variable.

Self-rated cost consciousness was divided into two groups using a median split. The range for the variable was '1' representing 'not concerned about cost' to '7' representing 'very concerned about cost.' No subjects rated themselves at '7.' The median was '2' giving rise to two possible groupings: 1 and 2-6, and 1-2 and 3-6. The Anova test was performed using both groupings and the results were virtually identical. The 1 and 2-6 grouping provided the most equal distribution between the two groups (17 subjects scored '1' and 23 subjects scored '2' through '6.'

The Anova test was performed with self-rated cost consciousness and experience level crossed with search question, on the dependent variable heuristic index. There were no significant effects for heuristic index. Thus, the result of the test that used the treatment variable to look at the effect of cost consciousness on heuristic index was verified using self-rated cost consciousness in place of the treatment variable.

Test of H2: High heuristic search approach will result in higher recall; low heuristic search approach will result in lower recall.

The variable for heuristic level is a normalized index created from four search process variables (new terms, citations viewed, commands issued, and search cycles). This hypothesis was tested in two ways. One was to test it with Anova, including experience level, treatment and search question as independent variables in the analysis. The other was to use a stepwise multiple regression, with recall as the dependent variable and heuristic index, experience level and cost consciousness treatment as the independent variables.

To test the hypothesis using Anova, heuristic index was recoded into high and low groups by dividing the scores at the zero point of the distribution for each question. The other independent variables were experience level, cost consciousness treatment and search question; the dependent variable was arcsine square-root transformed recall. However, transformed precision and total search cost were tested as well as dependent variables. This method of testing allows for identification of any interactions that may exist among the variables.

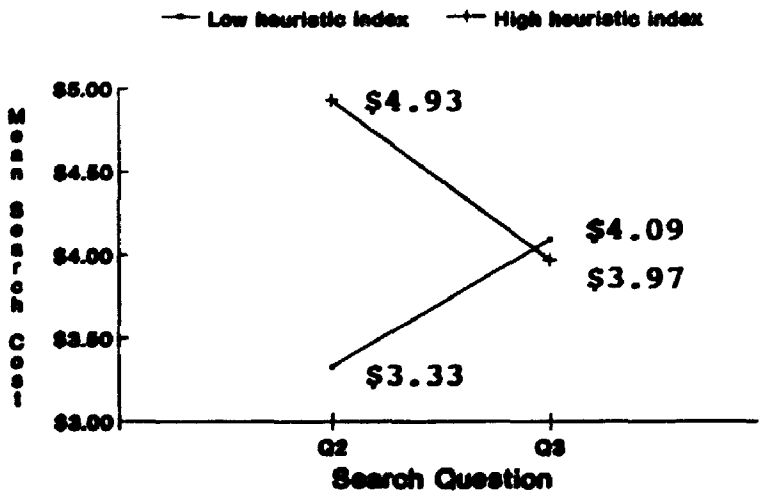
Did the high and low heuristic groups differ in terms of recall attained? For recall, there were no significant main effects or interactions on any of the independent

variables. However, it should be noted that the main effect of heuristic index did approach significance (F=3.63, p=.067).

When precision was taken as the dependent variable in the Anova, the main effect of search question was significant (F=5.36, p=.028). Mean precision for question 2 was 47.86 (SD=29.15) and for question 3 was 26.79 (SD=22.66).

In terms of total search cost, there was a significant main effect from cost consciousness treatment, as expected (F=9.03, p=.005). However, there was also an interaction between heuristic index and search question (F=4.58, p=.040). This is shown graphically in Figure 2 below.

FIGURE 2: Mean Search Cost for Groups Formed from Heuristic Index Versus Search Question



In terms of the hypothesis test using Anova, it is interesting to note that although total search costs and precision were shown to be significantly affected by the experimental design, recall was not affected significantly.

As mentioned above, the second method of testing this hypothesis was using multiple regression to look for a linear relationship between recall and heuristic index. Using this method, it is not necessary to recode the heuristic index variable, and thus lose information.

The first step was to do scatterplots of the recall versus heuristic index for each search question. For question 2, there was a great deal of scatter, but there was the appearance of a positive and somewhat linear relationship. However, for question 3, because there were only five relevant references in the answer set for the question, the recall variable becomes essentially a discrete distribution. This is obvious from the scatterplot, which shows distinct horizontal lines of points. However, discrete values for the dependent variable do not invalidate the multiple regression. Therefore, the test was performed separately for the two search questions.

The multiple regression tested the null hypothesis that there is no linear relationship between the variables, in this case between recall and heuristic index. This translates into testing that the slope of the regression line is zero, using the t statistic to determine

significance. The test was designed with the dependent variable recall, and independent variables heuristic index, experience level and cost consciousness treatment entered stepwise into the equation.

The result for question 2 was that the only significant variable entering the equation was heuristic index (t=2.098, p=.043, r-square=.14). This would indicate that the null hypothesis can be rejected and it can be assumed that there is a significant positive linear relationship between recall and heuristic index. This lends support for the hypothesis that a higher heuristic index will result in higher recall.

For question 3, no significant variables entered into the equation. The highest t-value was for heuristic index (t=1.449, p=.15).

Therefore, there does appear to be some support for hypothesis H2 that higher heuristic behaviour, as measured with the heuristic index variable, leads to higher recall. The relationship between the variables, however, is not strong. This is an interesting result, in that it tends to be an assumption among information specialists that more interaction with the system will improve recall. In this study recall was higher with greater heuristicity, but only for search question 2.

Test of H3: Extreme levels of topic familiarity (very high and very low) will result in a low heuristic search approach; moderate levels of topic familiarity will result in a high heuristic search approach.

Subjects were asked to rate their familiarity with the search topic on the pre-search questionnaire. A seven-point scale was used, with 1 representing 'not familiar' and 7 representing 'very familiar'. The Kruskal-Wallis test was used to look for differences in heuristic index scores between groups formed from self-rated familiarity with the search topic. The Kruskal-Wallis test requires that the dependent variable be at least ordinal level, but does not have to be normally distributed. It is the nonparametric equivalent of one-way analysis of variance. The nonparametric test was used because of the uneven distribution of scores across the seven groups formed for the topic familiarity variable. Under such circumstances, the Kruskal-Wallis test is more robust than oneway Anova. The tests were conducted separately for the two search questions and for the two experience groups.

For search question 2, there were no significant differences found among the self-rated topic familiarity groups for either experts ($X=7.6381$, $p=.18$) or novices ($X=6.0296$, $p=.30$). However, for question 3, a significant difference was found among groups for the experts ($X=11.5690$, $p=.02$). No difference among groups was found

for the novices for question 3 (X=2.8982, p=.57). The mean rank for heuristic index for each category of self-rated topic familiarity is given in Table 9 for experts searching question 3.

The uneven frequency distribution for the groups, makes data analysis difficult. The only clear conclusion is that for question 3, the expert group showed higher heuristic level in the middle ratings (4 and 5) for topic familiarity than the lower ratings (ratings 1 through 3).

TABLE 9: Mean Rank for Heuristic Index by Self-rated Familiarity with Search Topic (Question 3)

Topic Familiarity Category	*Mean Rank Heuristic Index*	
	N	*Experts*
1 (Not Famil.)	6	6.33
2	3	8.00
3	5	9.00
4	4	17.75
5	2	16.00
6	0	-
7 (Very Famil.)	0	-

An Anova test was done with topic familiarity crossed with search question as the independent variables, and heuristic index as the dependent variable. There were no significant main effects; however, the interaction between topic familiarity and search question approached

significance (F=2.39, p=.053). The means for the topic familiarity groups are given in Table 10, below.

The overall means for heuristic index are quite different for the two search questions: -.196 for question 2, and .090 for question 3. Given that there are seven groups formed for topic familiarity, the results are difficult to interpret. For search question 3, the 'very familiar' and 'not familiar' groups had the lowest values for heuristic index, and the intermediate familiarity groups had the highest values for heuristic index. This

TABLE 10: Mean Heuristic Index Score by Topic Familiarity Group, by Search Question

Topic Familiarity Category	Question 2 Mean (S.D.)	Question 3 Mean (S.D.)	N*
1 (Not Famil.)	.570 (.156)	-.570 (.552)	2
2	-.418 (.471)	-.380 (.457)	5
3	-.025 (.720)	.311 (.859)	8
4	.257 (.576)	.037 (.954)	3
5	-.483 (.655)	.194 (.834)	10
6	-.253 (.514)	.313 (.805)	7
7 (Very Famil.)	-.090 (.000)	-.460 (.000)	1
Sample Total	-.196 (.627)	.090 (.782)	36

* Note: One subject failed to complete question 2 and three subjects failed to complete question 3.

result lends support to the hypothesis. However, for question 2, the results cannot be so clearly interpreted, in that the 'not familiar' group had the highest mean

heuristic index, but the values for heuristic index tended to be lower for those having some familiarity with the topic.

It appears, not surprisingly, that the search question and self-rated topic familiarity interact to influence the heuristic index variable. It would be useful to know what qualities of the search question, in addition to topic familiarity, influence the heuristicity of the searcher.

Test of H4: Extreme levels of database familiarity (very high and very low) will result in a low heuristic search approach; moderate levels of database familiarity will result in a high heuristic search approach.

The measure made for database familiarity was to have subjects describe the proportion of searching done on the ERIC database. They were given four categories to select from: <10%, 10-50%, 51-80%, >80%. It was intended that the <10% category would include those with no ERIC experience; however, 4 subjects wrote in 0%. It is impossible to tell whether all subjects having no experience with ERIC wrote in 0%; therefore, the two categories, 0% and <10%, were combined into the 'low' experience category. Only two subjects had more than 80% of their searching done on ERIC.

Effect of familiarity with the ERIC database on heuristic index was tested using the four categories and

the Kruskal-Wallis test. The mean ranks for the categories are given in Table 11.

There were no significant differences between the groups on either search question, for either experience group. However, the distribution of cases over the groups is very weighted, with the <10% group having more than half the cases. This unequal distribution makes the result of the Kruskal-Wallis test suspect.

TABLE 11: Mean Rank for Heuristic Index by Self-rated Familiarity with ERIC Database

ERIC Database Familiarity Category	*Mean Rank Heuristic Index*			
	N	*Experts*	*N*	*Novices*
Question 2				
<10%	14	10.21	14	8.75
10-50%	4	9.25	3	10.14
51-80%	1	19.00	0	-
>80%	1	11.00	0	-
Question 3				
<10%	14	9.93	12	8.17
10-50%	4	10.50	3	7.33
51-80%	1	9.00	0	-
>80%	1	20.00	0	-

In addition to the measures described above for proportion of searching done on the ERIC database, subjects were also asked to rate themselves as having or not having experience with the ONTAP ERIC training database (a 'yes'

or 'no' response). An Anova test was conducted with ERIC experience crossed with search question as the independent variables, and heuristic index as the dependent variable. There was a significant main effect of question on heuristic index ($F=4.17$, $p=.049$). Question 2 had a mean heuristic index of -.178 ($SD=.627$), and question 3 had a mean of .087 ($SD=.794$). Therefore, there was a significantly higher mean heuristic index for question 3 than for question 2, although ERIC experience had no significant effect.

In summary, given the large number of subjects falling into the category of <10% of past searching having been done on the ERIC database, it is difficult to test this hypothesis. With such an uneven distribution for ERIC experience, the tests are suspect. In addition, the investigation of the effect of having any previous ERIC experience versus no previous experience suggests that ERIC experience is not a good predictor of heuristic index, whereas the search question itself has a significant effect.

Test of H5: Extreme levels of anxiety (very high and very low) will result in a low heuristic search approach; moderate levels of anxiety will result in a high heuristic search approach.

Anxiety was self-rated on a seven-point scale, with 1 representing 'not anxious' and 7 representing 'very anxious'. The rating was made at the completion of the online session for each search question. The Kruskal-Wallis test was used to test whether the groups were members of the same population. The mean ranks for the groups formed by self-rated anxiety are given in Table 12, for each search question.

The Anova test was performed with anxiety crossed with search question as the independent variables and heuristic index as the dependent variable. No significant main effects or interactions emerged between anxiety and question; however, the main effect of question approached significance ($F=3.46$, $p=.073$), with mean heuristic index being higher for question 3 than for question 2. Therefore, hypothesis H5 that extreme levels of anxiety will result in a low heuristic search approach is not supported.

TABLE 12: Mean Rank for Heuristic Index by Self-rated Anxiety

Anxiety Category	*Mean Rank Heuristic Index*			
	N	*Experts*	*N*	*Novices*
Question 2				
1 (not anxious)	3	12.33	4	7.13
2	4	13.50	4	8.50
3	4	10.50	1	12.00
4	2	9.00	2	10.75
5	5	7.20	5	8.40
6	2	11.50	2	16.50
7 (very anxious)	0	-	0	-
Question 3				
1 (not anxious)	2	10.50	1	7.00
2	4	6.00	4	7.63
3	1	9.00	3	9.67
4	2	16.50	3	7.00
5	5	14.40	3	5.83
6	6	8.50	2	16.00
7 (very anxious)	0	-	1	16.00

ADDITIONAL DATA ANALYSES

Relationship Among Search Outcome Variables

Recall and precision are often said to be inversely related, that as one rises the other falls. For the two search questions in this experiment this was not found to be true. For search question 2, the value of Pearson's r for recall and precision (untransformed) was .14, and for square root arcsine transformed recall and precision was

.22. For question 3, it was .61 for untransformed recall and precision, and was .64 for transformed recall and precision, both of which were significant at the .001 level. Therefore, for the third search question, which had an answer set of only five relevant references, there was a high positive correlation between recall and precision.

Direct Effect of Cost Consciousness on Recall and Precision

Hypothesis H1 looked at the effect of cost consciousness on heuristic index. Hypothesis H2 looked at the effect of heuristic index on recall. It is useful to look directly at the relationship between cost consciousness and recall, to establish whether some direct effect exists outside of the relationships hypothesized between cost consciousness, heuristic index and recall.

The Anova test was used to look for any effect of cost consciousness treatment, experience level and search question on recall and precision. When recall is taken as the dependent variable, there were no significant main effects or interactions. Thus cost consciousness does not appear to have a direct effect on recall. However, it should be noted that the interaction between the variables experience level and search question approached significance ($F=3.31$, $p=.078$). For search question 3, experts had higher mean recall for both the treatment and control conditions. However, for search question 2,

novices had higher recall than experts under the cost consciousness treatment, but had lower recall than experts under the control conditions.

In terms of precision, there was a significant main effect from the question variable (F=12.53, p=.001). The mean for precision on question 2 was 47.86 (SD=29.15) and on question 3 was 26.79 (SD=22.66). Therefore, precision was significantly higher for search question 2 than it was for question 3.

Relationships Among Self-rated Variables

It is useful to explore associations among the self-rated variables. Only the values of Pearson's r exceeding plus or minus .40 will be reported (see Appendix 6, Tables B1(a) and B1(b), for a complete report).

For question 2, topic familiarity was positively associated with confidence that the best search terms had been found (.52). Thus, the more familiar the topic, the more confidence in the search terms chosen. For question 3, topic familiarity was negatively associated with rating of lack of interaction with the requestor being detrimental to the search (-.51) and need for more requestor interaction to improve the search (-.43). In other words, the greater the familiarity with the topic, the less perceived need to interact with the person requesting the search. This would seem to indicate that subjects felt

that problems with understanding the topic were linked to not being able to talk to the person requesting the search.

Lack of interaction with the requestor being detrimental to the search was negatively associated with confidence that the best search terms had been chosen (-.51), for question 3. That is, the greater the need to interact with the requestor, the lower the confidence in the adequacy of the search terms chosen. For question 3, it was positively associated with need for more requestor interaction to improve the search (.74). This is a measure of the strength of association between the perceived need for interaction with the requestor before the search and the perceived need for interaction with the requestor following the search. It is interesting that the strength of association for the variables was significant at the .001 level for question 3, but was quite low for question 2 (.35). It would seem to indicate that for question 2 the interaction with the system somewhat alleviated the feeling of needing to talk with the requestor, whereas for question 3, the need remained relatively great after the search.

For question 3, there was a negative association with need for more interaction with the requestor and pre-search confidence that the best search terms had been chosen (-.46). It appears that, for question 3, a post-search feeling of greater need to talk with the requestor in order to improve the search is associated with a pre-search lack of confidence in the search terms chosen.

Post-search confidence that all relevant citations had been found was positively associated with satisfaction with the content of the answer set for question 3 (.70). In other words, the greater the satisfaction with the answer set, the greater the confidence in recall attained. It is interesting that this was not seen for question 2. This would seem to suggest that higher recall was a pricrity in assessing the goodness of the answer set for question 3.

Anxiety was positively associated with need for more interaction with the requestor (.41), for question 2 (that is, the greater the anxiety, the greater the feeling of needing to talk to the requestor). Surprisingly, this relationship did not emerge for search question 3. Once again, it is difficult to pinpoint the association between anxiety and other search variables.

In summary, it is interesting that no associations between self-rated variables held consistently for both search questions. It would seem that the self-ratings were strongly influenced by the nature of the search question.

Search Topic Familiarity and Search Performance

It was expected that degree of familiarity with the search topic would have an effect on search process and outcome. Pearson's r was used to look at the relationship between familiarity, a seven-point self-rated variable (1=very low, 7=very high), and the search process and

outcome variables. The values for the search process and outcome variables are given in Table B2, of Appendix 6.

For question 2, familiarity with the search topic was negatively associated with number of Select Steps commands issued (-.52) and number of modifications to search terms made in the course of the search (-.40), but was positively associated with number of Type commands issued (.41). In other words, higher familiarity with the search topic was related to fewer Select Steps commands, fewer modifications to search terms, and use of more Type commands. For question 3, familiarity with the topic was positively associated with number of descriptors used (.45), number of sets created (.49), and number of limits placed (.45; probably related to limiting search terms to descriptors). Thus, greater familiarity with the topic for question 3 was related to use of more descriptors, creation of more sets and placing more limits on the search.

Once again, it seems that there are considerable differences between the two search questions. For search question 2, it seems that higher familiarity with the topic was associated with use of fewer Select Steps commands (but not fewer Select commands) and with fewer modifications to the original search terms used. However, with greater familiarity with the topic there was a tendency to look at more citations. For question 3, greater familiarity with the topic was associated with use of more descriptors (and,

hence, more delimiters) and with creation of more search sets.

Effect of Experience, Question, and Cost Consciousness on Search Process Variables

The effects of experience and cost consciousness on search outcome and heuristic index were tested as part of the research hypotheses. Heuristic index is a composite measure derived from four of the search process variables: number of new terms introduced during the search, number of search cycles, number of commands issued, and number of citations viewed. It is useful to explore the effect of experience, cost consciousness and search question on each of the search process variables individually.

The search process variables measured in this study were selected to correspond to measures used traditionally in studies of online searching behaviour. This was done to allow comparisons to be made with past research. These variables attempt to quantify the various components of the actual terminal session, such as the number of each type of command issued and the number of citations viewed during the search. In this study, counts were made of number of Boolean ORs, ANDs and NOTs, number of search cycles, number of sets browsed (a set was browsed if at least one citation from that set was viewed), number of citations viewed, number of original search terms used from the pre-search

formulation, number of new or modified terms introduced during the search, total number of commands issued, and complexity of the final formulation (number of limits, parentheses, terms and operators employed to produced the set designated as the answer set).

The search process variables were tested for normality and found to be relatively normally distributed. Therefore, Anova was used to test the effects of experience level, cost consciousness treatment, and search question on each of search process variables. The results of the tests are given in Table 13.

The results suggest that experience significantly affects the number of Boolean ORs used in a search (experts used more ORs), the number of new search terms introduced (experts introduced more new terms), and the complexity of the search formulation leading to the final answer set (experts used more complex search formulations). Experience level was the only significant effect for the number of ORs and the number of new terms variables. Complexity, however, was also significantly affected by search question. Question 2 had a higher mean complexity than did question 3. This is not surprising given that question 2 had several distinct facets to be searched and combined.

TABLE 13: Anova Test of Search Process Variables: Treatment and Experience Level Crossed with Search Question

Effect	*F*	*Signif.*	*Means*		*S.D.*
Dependent Variable: Number of Boolean ORs					
Experience	9.49	.004	Experts:	6.05	5.10
			Novices:	2.43	1.98
Dependent Variable: Number of Boolean ANDs					
Treatment	5.82	.021	Treat:	2.20	1.11
			Control:	3.45	2.10
Dependent Variable: Number of Search Cycles					
Question	9.42	.004	Quest 2:	1.70	0.99
			Quest 3:	2.45	1.40
Dependent Variable: Number of Citations Viewed					
Question	4.22	.042	Quest 2:	7.30	5.66
			Quest 3:	10.88	11.18
Dependent Variable: Number of Sets Browsed					
Question	9.31	.044	Quest 2:	1.65	0.98
			Quest 3:	2.35	1.27
Dependent Variable: Number of Original Search Terms					
Experience by Question Interaction	4.32	.045	(See Figure 3)		
Dependent Variable: Number of New Search Terms					
Experience	9.57	.004	Experts:	3.65	2.81
			Novices:	1.65	1.14
Dependent Variable: Number of Search Commands					
No significant main effects or interactions					
Dependent Variable: Complexity of Final Formulation					
Experience	13.92	.001	Experts:	12.93	6.53
			Novices:	5.91	4.68
Question	4.85	.035	Quest 2:	12.08	11.04
			Quest 3:	7.53	5.76

It is interesting to note that treatment emerged as a significant main effect for only one variable: number of

Boolean ANDs. The control group used more ANDs than did the treatment group, which may give some indication of the effect of cost consciousness on search behaviour. The fact that search question did not emerge as a significant effect is interesting, given that question 2 had several facets to be combined, so that it would be expected that more ANDs would be used for question 2 than for question 3. This was not the case.

Search question did emerge as a significant main effect on several variables. There were significantly more search cycles for question 3 than for question 2. More citations were viewed and more sets were browsed for question 3 than for question 2. As mentioned above, search complexity was higher for question 2 than for question 3.

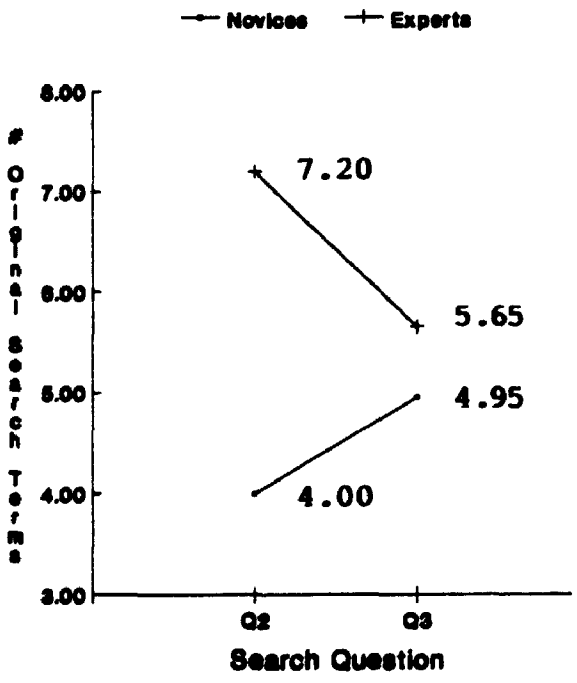
No significant main effects or interactions were found for total number of search commands. This is a problematic measure in that often the same result in a search can be accomplished by using one relatively complex command as using several commands. This measure has been used often in past research to describe search behaviour, yet it really does not tell us very much. It would be more useful to combine this measure with an indicator of complexity to get a more meaningful measure of the search effort involved.

Only one significant interaction was found. For the number of original search terms used (from the pre-search formulation), there was an interaction between experience

level and search question. This is illustrated in Figure 3 below.

Experts had more original search terms than did novices on both search questions. However, the difference is quite large for question 2, but there is a convergence for question 3. In other words, the mean number of original search terms for experts was higher for question 2 and lower for question 3, whereas the opposite occurred for the novices.

FIGURE 3: Mean Number of Original Search Terms for Groups Formed from Experience Level Versus Question



Exploration of Other Self-rated Variables

Subjects were asked to rate themselves on several additional variables thought to influence search performance. It is useful to look for associations that might exist between these self-rated variables and the search process, search outcome, and heuristic variables. Correlation was used to look for associations, and the results are tabulated in Appendix 6, Tables B2 through B8. The results for three of the self-rated variables, for which moderate correlations were seen, are described below.

Lack of Interaction with Requestor

Subjects were asked to rate the extent to which lack of interaction with the requestor was considered detrimental to their ability to search the question (1=not detrimental, 7=very detrimental). This information was gathered because of the experimental set-up in which the subject was given only a brief written description of the search query. Comments were made by subjects during the pilot test regarding the difficulty of searching under such artificial conditions. The rating was given before beginning the search. Pearson's r was used to look for correlations between degree to which lack of interaction with the requestor was rated detrimental to the search and

the search process, heuristic, and search outcome variables (reported in Table B4, Appendix 6).

For question 2, there was a negative association with precision (-.43). The greater the feeling that not being able to interact with the requestor was detrimental to the ability to search the question effectively, the lower the precision ratio for the search.

For question 3, there were several negative associations. Lack of interaction with the requestor was negatively associated with number of Select commands (-.44), number of spontaneously derived search terms (i.e., terms that were not from the pre-search formulation, the browsed citations, nor modifications to previously used terms) (-.43), and heuristic index (-.45). In other words, a greater perception that lack of requestor interaction was detrimental to the search was associated with fewer select commands, fewer spontaneously derived search terms, and lower heuristic index scores.

Low values on search process variables have frequently been explained in the online literature as being a result of cost consciousness. It appears from these results that lower heuristic search approach and depressed search process variables are associated with a greater perceived detriment of not being able to interact with the person requesting the search, which is frequently part of the experimental design in research in online searching.

However, once again, the effect seems to vary considerably with the search question.

Satisfaction with Content of Answer Set

Following their search, subjects were asked to rate their satisfaction with the answer set they had retrieved (1=not satisfied; 7=very satisfied). This question is similar to the question on their confidence in retrieving all relevant references, but also takes into account other entirely individual beliefs in what makes up a 'good' answer set. What was the association between their image of the goodness of what was retrieved and search performance? The association measures for the search outcome, heuristic, and search process variables are given in Table B7, in Appendix 6.

Only complexity was correlated with satisfaction with the content of the answer set, for question 3 (-.47). A higher confidence in the answer set was associated with a lower complexity of the final search formulation.

Accuracy of Answer Set Size Assessment

Another question deserving of exploration is, for those subjects having an estimate of number of relevant references, how accurate was that estimate? Subjects were asked, following each search, to assess the size of their

answer set. They were given a choice of four responses: unknown, too large, too small, and about right. The distribution of responses for the two questions is given in Table 14.

The distribution did not differ substantially for novices and experts. For question 2, both groups had the greatest frequency in the 'about right' category, and for question 3, both groups had the greatest frequency in the 'too small' category. It is interesting that novices resemble experts in their assessment of answer set size. This may be a characteristic that does not change much with experience. There were no substantial differences in the distribution of answer set size assessment for treatment and control groups.

TABLE 14: Frequency Distribution of Set Size Evaluations

Response	*Question 2*	*Question 3*
Unknown	1	1
Too large	3	4
Too small	13	21
About right	23	14

This measurement was made in an attempt to determine the accuracy the subjects' assessment of how well they had captured the relevant references in the database, without sacrificing precision too greatly. It called for an

impression on the part of the subject. Those assessing their answer as 'about right' presumably felt they had satisfied their own criteria for success. It could be argued that those answering 'too small' felt they had not attained high enough recall, and that those answering 'too large' felt they had sacrificed precision too greatly. The 'unknown' group was dropped from further analysis because it contained only one case for both search questions. One-way Anova was used to compare the three groups, set size 'too large,' 'too small' and 'about right,' in terms of recall and precision. Experts and novices were treated separately, under the assumption that overall search experience might lead to more accurate set size assessments. Testing was done separately for the two search questions.

There were no significant differences among groups on either recall or precision for either experience group. This was true for both search questions. Apparently, assessment of the goodness of an answer set by its size is not a good predictor of search outcome.

DISCUSSION

INTRODUCTION

In this section, a summary of the most important findings of this research is given. Following this is a more detailed discussion, placing these findings and their implications within the context of previous research into online searching. The relevance of this study is then assessed in the light of the emergence of CD-ROM technology, and end-user searching. Finally, questions raised for further research are discussed.

SUMMARY OF MAIN FINDINGS

One goal of this research was to look at the effect of searchers' cost consciousness on the search process, while controlling for differences in amount of search experience. Another goal was to determine how heuristic approach influences search success. An extension of this problem was to determine the factors that influence heuristic approach.

There was evidence that the cost consciousness treatment imposed on subjects in this study was effective in modifying both their search behaviour and their concern for overall search cost. Subjects receiving the cost consciousness treatment had significantly lower total

search costs than did the control group subjects. In addition, those receiving the treatment reported significantly greater concern for search cost, as determined by self-rating.

The search question itself was found to have a strong influence on search process and outcome. For many of the tests performed, the search question emerged as a significant effect. It is not surprising that search question should have such a strong effect; however, it highlights the importance of taking this variable into account in research into online searching.

An important finding of this research was that higher cost consciousness did not result in significantly lower scores for heuristic index, nor did higher cost consciousness result in significantly lower recall or precision. This was equally true for novices and experts.

It was hypothesized that higher values on the heuristic index variable would lead to higher recall. There was evidence of a positive relationship between these two variables; however, it is interesting to note that the strength of association is not great. What seemed an obvious relationship turned out not to be as straightforward as expected.

An important indicator found in this research was that there was a significant relationship between denying the searcher the opportunity to interview the person requesting the search and search performance. This effect was

greatest for search question 3, which was a topic that was unfamiliar to most of the searchers, and for which there was little information given within the statement of the query.

These findings are discussed in greater detail in the following section. The relevance of these results to previous research will be discussed as well.

DISCUSSION OF MAIN FINDINGS

Effect of Search Question

Searchers performed two search tasks and the nature of the questions searched had a significant effect on a number of variables. It is useful, therefore, to place the discussion of the results of this study within the context of the search tasks. The nature of the questions posed some problems in terms of analysis, but also occasioned some interesting results.

The two search questions used in this study differed significantly on a number of the search process variables. In terms of search process, there was a significant difference between the questions in the number of search cycles, number of citations viewed, number of sets browsed, and in the complexity of the final search formulation, with question 3 having significantly higher values on these variables. These variables can be interpreted as

indicators of attempts to gain feedback from the system to improve the search and to keep it on track.

There was no significant difference between the two questions in recall attained. However, precision was significantly higher for question 2 than for question 3.

It is interesting to note that the relationships among the self-rated variables were different for the two search questions. For question 2, topic familiarity was positively related to confidence in search terms chosen, and anxiety was positively related to need for more interaction with the requestor to improve the search. In other words, for question 2, the greater the familiarity with the topic of the search, the greater the confidence placed in the search terms, and the more anxiety felt, the stronger the feeling of needing to talk to the requestor. These relations were not seen for question 3.

For question 3, the greater the familiarity with the search topic, the lower the feeling of needing to talk to the person requesting the search. It appeared that, for question 3, subjects may have associated problems with understanding the topic with the inability to interact with the person requesting the search. It is interesting that this relationship did not emerge for question 2. The greater the need for requestor interaction, the lower the confidence in the search terms. There was a strong positive relationship between pre-search and post-search feeling of needing to talk to the requestor of improve the

search. This may be taken as an indicator that for question 3 the interaction with the system did not tend to lessen the need to talk to the requestor. This relationship was not seen for question 2.

Another question-based difference was found for the heuristic index variable when question was crossed with differing levels of ERIC database experience. There was a significantly higher mean heuristic index for question 3 than for question 2, although ERIC experience did not emerge as a significant effect. It appears that searchers were more highly interactive for question 3 than for question 2, when experience with the database was controlled for in the analysis.

It is necessary to analyze the nature of the two search questions to understand the question-based differences observed. Question 2 was a topic familiar to all subjects because it had a library science basis, and dealt with offering library services to the handicapped. It was a multifaceted query with a number of possible synonyms for each facet. Search question 3 (white flight to the suburbs), however, could be searched effectively by simply entering the expression **white(w)flight** or **white and flight**. Either of these approaches would result in a set having four out of the five relevant references in the database. This explains the high correlation between recall and precision on this search question. The ERIC thesaurus used in the experimenc instructs the searcher,

under the term **white**, to use instead **caucasian**. However, if **caucasian** is coordinated with **flight**, or any supposed synonyms, such as **migration**, **movement** or **relocation**, there are no hits. The searcher must then look for a way to not only improve the search but to find any hits at all. Introducing the term **suburb**, particularly as the initial search term, led many subjects to disaster. None of the relevant references used the term **suburb**, which is implied in the concept of white flight.

The discussion of question-based differences in search behaviour cannot be separated from the concept of familiarity with the search topic. Topic familiarity was a self-rated variable. The basis of the hypothesis about topic familiarity was derived from the theories of Streufert and Streufert (1978), which suggest that there is an optimal level of incongruity in the environment and moving beyond that optimal level in either direction (too much or too little incongruity) will lead to reduced information search. Applying this theory to the problem solving task of online searching, it was expected that being very unfamiliar with a topic, the searcher would be at a loss as to how to proceed. Similarly, being highly familiar with a topic, the searcher would tend to type in all the best terms and take the result without much further interaction, i.e., once the best terms have been used, there is nowhere else to go. However, knowing something about the topic, but not necessarily enough to formulate a

search with all the best terms, the searcher would try to learn more about the topic and its treatment in the database through interacting with the system: learning how the topic is indexed, learning the extent of its coverage in the database, learning about related and peripheral areas of the topic.

Although the hypothesis that moderate levels of topic familiarity would lead to greater heuristic index values was not supported, some interesting differences emerged for the two search questions. In fact, an interaction was observed between topic familiarity and search question, with a pattern similar to the one hypothesized appearing for question 3.

For the second search question, which was a generally familiar search topic, there was no difference among the groups having differing levels of topic familiarity for either experts or novices. However, for search question 3, a different result was observed. Because of the general unfamiliarity of the search topic, no 'very familiar' (rating of 6 or 7) group emerged, so that the test had to be conducted on the 'not familiar' through 'moderately familiar' groups. No differences emerged for the novices, but a significant difference did emerge for the expert searchers. Those rating themselves as moderately familiar with the question had a higher mean rank on the heuristic index variable. Because none of the searchers rated themselves as very familiar with the topic, it is

impossible to know how heuristic approach would have been affected for those very familiar with the topic. However, expert searchers who were very unfamiliar with the topic scored lower on heuristic index than did those who were at least moderately familiar with the topic. It is interesting that a similar result was not observed for novice searchers.

In searching question 3, "white flight to the suburbs," a great many searchers went online having no idea or only the vaguest idea of what the question was about. Some began the search by ruling out possibilities, e.g., searching **birds and suburbs**, thinking that the topic might be about migrating birds. Others tried to find a clue from an initial exploratory search as to what the topic was about, and there were a great many 'ah hahs' and 'so that's what this is about' early on in the search. This would be followed by a quick strategy reformulation or confirmation of a chosen strategy. Thus, for question 3, interacting with the system was essential for most searchers for clarifying the topic and identifying the indexing used in the database.

For search question 2, however, the topic was very clear before the searchers sat down at the terminal and the difficulties with this question arose in defining the extent of the topic and in finding the best indexing. There were a few peculiarities in the indexing of the topic that resulted from ONTAP ERIC being a small subset of ERIC,

e.g., **physically handicapped** is a descriptor in the ERIC thesaurus, but only retrieves one citation in the ONTAP database. For this question, most subjects went into the search thinking their problem would be to narrow the retrieval set to a manageable size, when in fact, the problem became one of expanding each of the facets so as to be large enough for an overall intersection. When the primary facets of the search were formulated too specifically and intersected, the result was usually zero, causing initial exclamations of surprise and confusion, and need for a rapid readjustment of the search strategy.

To extend and improve the study of effects on search performance of topic familiarity and differences across search questions in, it would be necessary to look at a wider range of questions. This could be done in an experimental or natural setting by having searchers rate their familiarity with the topic before beginning the online session, and after completing the preliminary work in formulating the search. It is difficult to draw any conclusions looking at only two search questions, but it would be worthwhile to pursue this further to determine the effect of topic familiarity and other question-based variables on heuristic approach.

In summary, it is apparent from this study that large differences in search behaviour are related to aspects of the search question, which is not surprising. It seems that the nature of the question, particularly familiarity

with the topic, stimulates a need for quite different types of information, for instance, help in determining the best search terms is not so important for a familiar search topic.

The search questions were categorized by ONTAP ERIC as being of intermediate difficulty. However, it was apparent that subjects tended to find question 3 to be a more difficult search topic. It is certainly of interest to compare search performance for the two questions; however, extrapolation beyond the bounds of the experiment regarding question differences should be done with caution. Future research should include a variety of questions having very different features, to look for commonalities of behaviour across questions and to identify the question variables strongly influencing search performance.

It is clear that subjects performed quite differently on the two questions, which suggests that controlled research should take into account question difficulty in establishing search tasks. Fidel (1985) recommended that search tasks be kept simple and straightforward, citing a finding of high variability in the specificity with which search requests are interpreted. However, essential behavioural differences among searchers may only truly arise on challenging questions. In any future research investigating problem solving behaviour, it would seem essential to set search tasks challenging enough to ensure that searchers respond to the problem as non-trivial.

Cost Consciousness

There has been a great deal of speculation within the literature of online searching (e.g., Wanger et al., 1972; Fenichel, 1981; Vigil, 1983; Harter, 1984b) on reasons for low levels of interactiveness observed between searchers and the online system. A frequent suggestion has been that this may be a result of concern over keeping search costs down. However, this explanation had never been tested experimentally. One of the goals of this research was to test the effect of cost consciousness under controlled conditions.

Higher cost consciousness did not result in significantly lower heuristic approach, as measured with the heuristic index variable. Therefore, the first hypothesis of this study was not supported. The theory behind this hypothesis was that in an effort to keep search costs down the searcher would minimize interaction with the system, resulting in a lower heuristic index score, and would rely less on feedback obtained from the system to improve the search. The second research hypothesis suggested that lacking valuable information and feedback from the system, necessary for better problem solving, the search success would suffer and lower recall would result. This was, in fact, observed; greater heuristic approach was found to be related to greater recall. Therefore, no

indirect effect of cost consciousness on recall was observed, nor was there evidence of a direct effect of cost consciousness on recall.

The lack of support for the first hypothesis is a rather interesting finding. Cost consciousness has come to be a useful explanation for any somewhat counter-intuitive observations of searcher behaviour. Apparently the effect of cost concerns is not a straightforward phenomenon. Several explanations may be possible for this result.

If it is assumed that the measure for heuristic approach (heuristic index) is valid and measures what it purports to measure, then it appears that cost concerns may affect the search process but not significantly affect the outcome of the search, at least for the search tasks used in this study. The heuristic index measure is based on four traditional measures of search process that are generally taken as measures of search effort: number of search cycles, number of commands issued, number of citations viewed, and number of new terms introduced in the course of the search. Higher values on these measures result in a higher value for the heuristic index.

It is possible that the artificial search situation created for this experiment altered the behaviour of the subjects. However, both treatment and control groups were under the same constraints of working in an artificial situation. Presumably, then, given that the treatment did

have a demonstrable effect, some differences in the dependent variables should have emerged, if they existed.

This study had the interesting result of searchers lowering their search costs without significantly reducing their search success. In other words, they became more efficient searchers. What was perhaps more surprising was that this was true for novices as well as experts.

The attitudes of the experts towards the treatment is interesting. The experts repeatedly commented that ERIC was a cheap database and that ONTAP ERIC was extremely cheap. It would seem likely that they were less impressed by the cost consciousness treatment for these databases than they would have been for a very expensive database. It was quite clear from comments made by the searchers that the experts were assessing costs during the experiment relative to costs they characteristically incurred in their work situations. Novices did not have the benefit of this sort of comparison due to lack of experience and so tended to make numerous comments about 'spending all your money', and 'this is costing a lot', and so forth. Nevertheless, there was no significant difference found between experts and novices in the effect of cost consciousness on heuristic index or recall.

It is important that cost consciousness behaviour be studied on the more expensive databases to get a true picture of the phenomenon. This is especially true for expert searchers. It would be erroneous to extrapolate too

far beyond the bounds of this research as to the effects of cost consciousness on search performance. It is probably safe to say that they do modify their behaviour to a certain extent, and do tend to reduce their heuristic approach for some questions. However, it would be only on the expensive databases that the experts work with from day to day that cost constraint effects could be effectively tested.

In summary, it appears that within the parameters and constraints of this study, both experts and novices were able to respond to a cost constraint without a significant reduction in heuristic approach or degradation of search performance. It would be necessary to test this on a wider range of search questions and databases to verify this result.

Heuristic Index and Recall

Searchers in this study were given as their goal to find an answer set having all citations relevant to the question, i.e., maximum recall. They were given no instructions regarding precision levels to aim for. It was hypothesized that a high score for heuristic index would relate to greater search success, in this case defined as higher recall. It has been a basic assumption in the area of online searching that the greater the interaction between the searcher and the system, the better the result

of the search. This seems a logical assumption. However, the nature of the interaction must be taken into account; for instance, simply being online for a longer time does not mean that a better search should result from the interaction. Greater heuristicity assumes certain kinds of actions, particularly those that help the searcher gain feedback from the system to improve the search and keep it on track. In this study, this is referred to as a higher heuristic approach, as measured with the heuristic index.

There was evidence from this study that a higher score on the heuristic index variable was related to higher recall. This relationship approached but did not attain significance with the Anova test. When the relationship was tested with multiple regression, a significant positive association was found.

An interesting finding related to this relationship was that experience level of the searcher did not appear to significantly affect the association between heuristic index and recall. Past studies have reported that novices perform surprisingly well in terms of recall attained, often as well as the expert searchers. In this study, even under a cost constraint, novices maintained a performance level somewhat equivalent to that of the experts.

The essential research question underlying this hypothesis was whether being more interactive with an online system, and obtaining more feedback from the system, would result in more successful searching. The difficulty

is in discerning the nature of that interactiveness. It is possible to issue many commands, to use many search terms, and to look at many retrieved citations, but essentially absorb little feedback from the system that is useful in solving the problem.

One important point for consideration is the extent to which the searcher is receptive to the feedback from the system. In several instances in this study, when searching question 3, the term "white flight" appeared in the title of the citations that were viewed on the screen; however, the searcher failed to see it or to use it to modify the search. In some cases, the searcher was attending to different information, particularly the descriptors that had been assigned to the record. Searchers often seemed to be overly reliant upon or to place excessive emphasis upon descriptors. In the case of question 3, this had a negative impact upon recall, where effective searching depended upon use of free text terms.

Other researchers have pointed to this phenomenon of failing to attend to relevant information. Fidel (1987) pointed out the difficulty for novice searchers to follow new leads in a search. Marchionini (1989) found that the younger children in his study failed to extract relevant information from retrieved text. This may be a deficiency in problem solving behaviour that reflects a tendency to focus too greatly on a sub-goal, losing track of the final goal.

To summarize, the relationship between search success and heuristic approach, that is, the process of interacting with the system to obtain information that would be useful in performing a better search, is difficult to quantify. Measuring heuristic approach was a crucial problem in this study. Attempts at obtaining an impressionistic rating by independent coders resulted in too much variability and an unreliable measure. The heuristic index measure was developed to quantify interactiveness with the system, using search process measures that individually had some inherent potential as measures of attempts to gain feedback from the system. Unfortunately, it is difficult to know what use is being made of the information that is obtained, whether searchers are attending to feedback that is useful or useless in providing information to assist in strategy reformulation. For example, one searcher in this study viewed numerous citations but focussed almost exclusively on the age of each citation and paid little attention to the subject relevance or the indexing used. In other words, the person was obtaining feedback that was virtually useless for improving the search.

The individual variability of people as information processors will always be problematic in studies of problem solving behaviour. Nevertheless, the heuristic index measure seemed to be sufficient at a coarse-grained level in giving some quantification to heuristic approach. Certainly, more work is needed to find a measure for

heuristic approach of online searchers that will encompass all aspects of this important phenomenon.

Denying the Searcher Interaction with the Requestor

The experimental setup did not reflect the real world situation in that subjects were given only a brief written description of the search question and were unable to interact with the requestor to clarify the topic. This method has been used in the past by other researchers because it provides a method of control over the experimental situation, and for other purely practical reasons.

Given that it was an artificial situation, subjects were asked two questions aimed at discovering the degree to which this situation influenced them. Before beginning the search, they were asked to rate the extent to which they felt that lack of interaction with the requestor was detrimental to their search. After completing the search, they were asked to rate the extent to which they felt that the search could have been improved had they been able to interact with the requestor. The results indicate that, for one of the search tasks, search question 3, the greater the perceived negative impact of not being able to interact with the requestor, the lower the measure for heuristic index.

The finding that search behaviour is affected by not being able to interact with the requestor is commensurate with those of other researchers. Various experiments have been conducted with subjects being given only a written description of the search topic (e.g., Fenichel, 1981). Fidel (1985) commented on the frustration expressed by searchers who were given only a written statement of the query, which was also found in the current study. It would seem that care should be taken in interpreting results from a research design that does not allow interaction between the requestor and searcher. It is especially difficult to interpret results of designs in which some subjects searched a real request and others were forced to search from a written query statement (e.g., Harris, 1986).

Salomon and Burgess (1984) surveyed university librarians to determine their attitude toward patron presence during a search. They found that 77% of the searchers agreed that they prefer the patron to be present, and that the most important factor in wanting the patron present was as an aid to better search refinement.

In an observational study of human-computer interaction using an online search system, Baker and Eason (1981) found that 31% of elapsed time was devoted to searcher-requestor interaction. They found an average of 34 conversations per session, of which 66% were less than 10 seconds in duration. This indicates that at many points

in the search session, the searcher and requestor felt the need to interact.

Saracevic and Kantor (1988b) found higher recall scores for the group of subjects having a taped statement of the problem and intent of the requestor. The lowest recall was from the group having written search terms removed from the context of the problem statement. Again, this would tend to reaffirm the negative impact on the search process of denying the searcher an opportunity to interact with the requestor.

Cost consciousness did not emerge as a significant factor affecting the heuristic index score, nor did experience level significantly affect this variable. The question remains, then, as to what does affect heuristic approach.

It may be that there is a rather nebulous factor in searching that involves the extent to which the searcher makes the search query his or her own problem. The acceptance of the problem as a personal one may stimulate more information seeking, and a greater desire to learn from the system about the query topic. As Kuhlthau et al. (1988) noted in their analysis of cognitive and affective aspects of the information search process, "the critical point of the search process -- the turning point when the subjects shifted from uncertainty to confidence -- frequently was associated with forming a focus or a personal point of view about a topic. This was considered

evidence of cognitive movement toward sense-making" (p.70). It is possible that a significant factor in this 'internalizing' of the problem results from the interaction with the requestor, and that some searchers are strongly inhibited by not being able to interact with the requestor, while others seem to teach themselves the subject as they search. This seemed to be particularly true in this study for the 'white flight' question.

In terms of search outcome, there was a negative association between lack of requestor interaction being detrimental to the search and precision, for search question 2. This was reflected in comments made by the searchers on this question that they simply did not know how to limit the search without talking to the requestor. There was a great deal of difficulty with question 2 in defining the specificity of the topic and many subjects commented that it was difficult to know when to stop adding terms for various disabilities. One expert searcher commented, "Well, you know, this is ERIC and you could go on adding terms forever."

It was clear from the transcripts that subjects had some difficulty with the search situation and very often expressed doubt about their ability to search effectively without talking to the person requesting the search. They would often begin the search with comments about what they would have asked the person requesting the search and how that person could have clarified the topic for them. For

search question 2, "Library services to the physically handicapped (not mentally or language handicapped)," they often expressed uncertainty about the range of the topic and difficulty with deciding when to stop adding terms for handicaps. For search question 3, "White flight to the suburbs," the comment heard most often was that the topic was completely incomprehensible to them and if they had been able to talk to the requestor first, they would at least have known what it was about, and would have been able to get some search terms to begin the search.

A large number of subjects, both experts and novices, took pains to point out that their answer set would not be the true final answer. It was often stated at the conclusion of a search that what would happen at that point was that the requestor would be given the final set and together with the searcher would refine the search based upon the index terms in the citations. Subjects were often reluctant to tag a set as the final set. One expert subject gave the number of the answer set but stated "I am not finished". The inability to perform iterative searches is a definite problem with the experimental situation and hindrance to studying online searching as a true problem solving situation.

The findings of this research highlight the limitations of studying the search process using this methodology. Search performance may be more constrained than in the 'real world' where searchers would normally

interview the requestor prior to the search and would often go through several iterations of talking to the requestor and modifying the search before a final result is obtained. Subjects in this research frequently stated that they were not finished the search with the answer set given as the final one. They stated that 'normally' they would give the answer set to the requestor and redo the search after discussing it with the person.

Caution should be used in attributing reasons for lack of interactiveness in searching, such as cost consciousness, when the experimental design does not permit interaction between the searcher and the person requesting the search. It is clear from this research that many searchers feel themselves to be at a strong disadvantage in trying to search a question 'blindly.'

OTHER FINDINGS

Novice-Expert Differences

Some interesting information can be gleaned by looking at effects of experience on search behaviour. Novices used fewer Boolean ORs than experts while searching, as well as fewer terms entered into the search from the pre-search formulation, and fewer new terms added in the course of the search. Novices made more errors in syntax and in applying Boolean logic, a finding similar to that of other

researchers (e.g., Fenichel, 1981). Experts in this study scored significantly higher on complexity of the final search formulation than did novices, reaffirming the findings of Harris (1986).

It is quite apparent that novices fail to use the Boolean OR operator while searching. This reaffirms the identical finding by Sewell and Teitelbaum (1986) of end-users searching MEDLINE. Novices were weak in their ability or inclination to add synonyms and term combinations while searching. This is probably an area where online instructors should concentrate in getting novices more familiar with Boolean logic and with combining terms and sets. This skill is far superior in expert searchers, who apply Boolean logic with little apparent cognitive strain, and few errors. Much of the search time and increased costs incurred by novices seemed to result from puzzling over Boolean logic, or correcting errors in Boolean logic.

Often novices displayed what Blair (1980) referred to as an anchoring phenomenon on this question. They would keep a set representing one facet, e.g., library services, constant and keep ANDing different sets with it until they could get a reasonably sized result. They would also cling to the **library services** term (which was taken directly from the written problem st tement), even though combining it with the handicap facet would result in zero retrieval. These observations parallel those of Harter (1990), that

novices demonstrated poor ability to identify elementary sets that were contributing to poor recall. Experts would fairly quickly assess the situation, recalling that the database was small and fairly old, and as such, the indexing may have changed considerably. They would, for instance, quickly drop the qualifiers and expand the facets (**physical handicaps** to **handicap*** and **library services** to **librar***) to get any citations at all, and start looking at the retrieved citations for clues to indexing used. It appeared to be far more difficult for novices to do such a reformulation.

In this study, some searchers, both experts and novices, were able to perform effectively with minimal knowledge of the search topic and no interaction with the requestor. A next step, it would seem, would be to study in more detail the behaviour of these successful problem solvers to identify 'ideal' behaviour in transcending the constraints of the problem. Novice searchers exhibited many of the deficiencies in problem solving described by Dorner (1983). It was apparent that novices had considerable difficulty with Boolean logic and with managing large numbers of sets. The anchoring phenomenon described by Blair (1980) was frequently observed in the searching of novices, particularly when set or term combinations retrieved unexpected numbers of references (too large or too small). The behaviour of both experts and novices deserves further study with respect to deficiencies in

problem solving, and would be expected to yield information useful for training and for designing better interfaces for online systems.

Confidence in Chosen Search Terms

An important aspect of problem solving is estimating the likelihood of occurrence of events under conditions of uncertainty. To explore this aspect of problem solving behaviour in online searching, subjects were asked to rate their confidence in the search terms they had chosen prior to the online session. From this, it was possible to examine the relationship between confidence in search terms and search performance.

There were no strong correlations found in this study for either search question between subjects' self-rated confidence in their chosen search terms and any of the search process or search outcome variables. It appears that searchers may not be very good at making such an estimate, particularly when searching without any information obtained from talking to the requestor that could aid in strategy formulation.

It should be kept in mind that subjects were asked to rate their confidence that they had found the best search terms before sitting down to do the search at the terminal. Often after beginning their search, subjects were surprised to find how far off track they had been in their pre-search

formulations, which might explain the lack of associations with search success. In other words, for many subjects their estimate of the 'goodness' of their search terms was poor, and therefore bore little relationship to search performance.

Assessment of Search Results

Another relationship that was explored in this research was how the searcher's satisfaction with the search result related to search outcome. Subjects were asked to rate their confidence that all relevant references had been found and their satisfaction with the content of their answer set. They were also asked to judge the adequacy of the size of their answer set.

It appears that neither novices nor experts were very good at assessing the success of their searching. Their ratings of confidence in having found all relevant references and their satisfaction with the content of their answer sets were not strongly correlated with the search outcome measures.

Their assessment of the adequacy of the size of their answer sets showed some differences between the two search questions. Most searchers, for question 2, thought their answer set was about the right size. For question 3, most searchers thought their answer set to be too small. However, there was no significant difference among the

different rating groups in terms of either recall or precision.

Assessment of adequacy of set size is an interesting factor given what is known about searchers' behaviour with respect to size of retrieved set. Blair (1980) described the futility point criterion, the largest set size that a searcher will be willing to browse, which would necessarily influence a measure such as heuristic index.

Observations of searchers' behaviour in this study also led to questions about the effect of minimum set size that a searcher would browse. Many searchers kept trying to increase the size of a small set that had been retrieved in response to a formulation that represented all facets in a query, without browsing the set first. This would suggest a deficiency in problem solving related to focussing inordinately on a sub-goal or sub-theme, i.e., trying to obtain a set size meeting some internal criterion or sub-goal rather than browsing the set to assess the contents.

There was also a tendency among some novice searchers to browse building block sets, before they were intersected, seemingly to make relevance judgements. This is a rather futile strategy, and a waste of time. It, again, suggests a tendency to focus too much on a sub-goal, losing sight of the final goal.

There also appears to be little difference between experts and novices in the ability, or lack of ability, to

assess the search result. It should be kept in mind, however, that almost all of the subjects in this study should be considered relatively inexperienced on this database. It would seem likely that accurate self-assessments of search success would increase with practice on and familiarity with a particular database.

Observations of Search Behaviour

It was apparent from comments made by subjects that they were continually having to remind themselves that ONTAP ERIC differs from ERIC and that it is a very small database. The low retrieval on descriptors was a source of consternation among all the subjects. They frequently reminded themselves of the age of the database and that they shculd be using terminology appropriate to that time. Subjects also, interestingly, seemed to try to 'place' themselves in that time and analyze the search questions in terms of that time period. For instance, for question 2, they would try to orient themselves in terms of what was happening in libraries in 1975, to recall the date of the International Year of the Disabled, and so forth. For question 3, they would use the year 1975 to try to orient themselves in terms of racial conflict and busing. They would also frequently point out that this was an American database and so, although 'white flight' was something foreign to Canada, it was undoubtedly important in the U.S.

at that time. One of the novice subjects seemed fixed on the age of the retrieved references in evaluating relevance, making comments to the effect that they would be too old to be relevant to the requestor, apparently forgetting that all the references in the ONTAP database were from 1975. This could be a manifestation of the deficiency in problem solving described by Dorner (1983) as encapsulation: getting caught up in a sub-theme and forgetting the overall problem.

Most subjects retrieved too few references with their first formulation, often none at all. They often had trouble reformulating their strategy to do a broader search. If they had prepared an alternative strategy, it was usually designed to narrow the search, anticipating retrieving too many references. It may be that searchers take a more algorithmic approach to a search and stick to it (using their pre-search formulation) as long as nothing unexpected happens, e.g., zero hits, too large or too small sets retrieved, too many false drops for a term or term combination. At that point they must become more heuristic in approach, which is a pressure point in a search (Bates, 1981). When search terms do not work out as expected the result can be surprise and confusion, resulting from a tendency to be overconfident in the likelihood of an event occurring, as described in decision theory. Many subjects said they would logoff and rethink their strategy when such a surprising event occurred.

Most subjects pointed out that they would go to the requestor for evaluation and feedback of the retrieved citations, and would possibly redo the search at that time. Morris et al. (1989) referred to a tendency to search, logoff and reformulate to save money. Other researchers have also indicated that this behaviour may be a result of cost concerns. It was probable, however, that a more quick-and-dirty approach was a result of (1) an experimental design that did not permit interaction with the person needing the information, and (2) a restriction to observing what amounts to the first iteration of a search for information. It is difficult to isolate and study one portion of the problem solving process, and get a true picture of the process.

There was a strong tendency to use descriptors. For question 3, there seemed to be a great reluctance to use the term given to describe the phenomenon in question: white flight. It is not clear whether this is a result of previous training, natural search style, or whether it was an artifact of the experiment (i.e., because they were given a thesaurus, they may have felt they were supposed to rely on it).

Expert searchers were much better at recalling the syntax of the command language, not surprisingly, although this was true even for those who rarely used DIALOG. They were also more competent at manipulating large numbers of retrieved sets. They were quicker than novices at scanning

sets and remembering how they were formed, for reuse in another formulation. Experts made few errors in Boolean logic and if they made an error it was almost always recognized immediately and corrected. They were also better at monitoring a search to keep it on track. Experts were able to tell when a wrong set or term had been used, back-track, and reformulate a whole series of steps with apparently little cognitive strain.

The results of this research cannot really be compared to other, more extensive research into the influence of database experience on searching (e.g., Fenichel, 1980; Howard, 1982) where the experimental design included finely controlled for differences in overall familiarity with ERIC. However, it should be noted that Fenichel (1980) found that subjects experienced with ERIC used more thesaurus terms than those lacking experience, which may relate to the finding here that number of original search terms was greater for the experienced groups. Howard (1982) concluded from her research that level of experience and database familiarity would contribute to better performance for relatively difficult searches. The results of this research do not indicate substantial or consistent differences in effects of database familiarity for questions of varied difficulty; however, testing was done under the constraint of all subjects being relatively unfamiliar with ERIC.

One difficulty with this study was in effectively measuring anxiety. Anxiety can be produced by many things in the course of a search, and very probably is increased by the experimental situation. This research did not investigate the effects of anxiety levels varying throughout the search. The subjects were asked to rate their anxiety after the search was completed and the basis for their rating is unknown, e.g., anxiety level at the end of the search, at the beginning of the search, somewhat averaged over the search, or compared to the previous search. It would be difficult, if not impossible, to isolate and analyze the effects of different sources of anxiety on the search process. Clearly, though, it is not necessarily a bad thing for a searcher to feel some anxiety while performing a search, and in fact, may be a good thing. As Kuhlthau et al. (1988) pointed out, personal construct theory (which describes learning as a process of construct building, involving both intellect and emotions) suggests that anxiety is a natural part of the information seeking process.

RELEVANCE OF THIS RESEARCH TO EMERGING TRENDS IN INFORMATION RETRIEVAL: CD-ROM AND END-USER SEARCHING

This study was designed to identify variables influencing the performance of online searchers acting as search intermediaries, in a controlled experimental

setting. One of the basic goals for this research was to look at the effect of cost consciousness on the online search process and search outcome. The question is of interest, in particular, given the emergence of CD-ROM technology and end-user searching in and out of the library setting.

Various studies and reports have emerged in the literature assessing the impact of CD-ROM technology on online services in the library (for example, Anders and Jackson, 1988; Halperin and Renfro, 1988; Harter and Jackson, 1988; and Brahmi, 1988). Brahmi (1988) assessed the effect of CD-ROM on end-user and mediated searching in an academic library. The conclusion was that "CD-ROM has had a dramatic lowering effect on online end user searching and has had little effect on mediated searching" (p.47).

It appears that CD-ROM is not replacing mediated searching, but offers a cost-effective alternative, in academic libraries, to end-user online searching. Anders and Jackson (1988) note that of four databases studied over time, "in the case of all four databases, the users who migrated in the largest numbers from online to laserdisk products were the users of the BRS/After Dark and DIALOG's Knowledge Index services. The people who had mediated searches conducted seemed to still prefer mediated searches, albeit in smaller numbers" (p.28). Halperin and Renfro (1988) also looked at the impact of the new technology on an academic library and concluded that

laserdisks, commercial, and local time-sharing are more complementary than competing technologies. Libraries can justify having all three formats available" (p.42).

It has been suggested that CD-ROM represents a useful stepping stone to end-user online searching, permitting search techniques to be learned without a cost pressure (Halperin and Renfro, 1988; Anders and Jackson, 1988). In addition, CD-ROM users tend to adopt the technology and migrate to online with a concomitant overall increase in demand for online services. Added to this is a general trend towards end-user searching as a result of a number of factors. Marshall (1990) describes some of these factors. There is a general trend towards marketing online services to end-users, coupled with a societal tendency toward self-service, which is increasing the number of people doing their own online searching. "Recent trends in hardware and software are making it easier for professionals to access online information systems on their own. These trends include: (1) greater availability of microcomputers in home and office settings; (2) increasing levels of computer literacy among professionals; and (3) development of new databases and interface software oriented to the needs of the end user" (Marshall, 1990, p.56).

It appears, then, that CD-ROM is primarily a technology complementary to online searching, and that the market for mediated searches is not diminishing. In addition, more and more end-users are doing their own

online searching. Therefore, studies of the behaviour of novice searchers, of the factors influencing search success, and of the effect of cost consciousness on search performance are as necessary today as they were a decade ago. It could be argued that studies of novice searcher behaviour are more critical now than ever if systems are to be designed that offer end-users effective access to large online databases.

QUESTIONS ARISING FROM THIS RESEARCH

Several questions arise from this research that deserve further study. These result from problems with methodology, inconclusive results for hypotheses, and some unexpected findings.

In terms of methodology, to truly look at the relationship between heuristic approach and search performance it is essential to include the interaction with the requestor. This is a crucial information gathering device that seriously affects information gathering behaviour during the online session. The extent to which heuristic approach is altered varies with both the difficulty of the search topic and with the individual searcher. Some are apparently more affected than others.

Searchers clearly had more difficulty with one of the experimental search questions than with the other. This raises some interesting questions about what characterizes

a difficult search. Further research could be done varying the characteristics of the search question and determining effect on search performance and perceived difficulty. For instance, the two questions in this research differed in familiarity of the topics, in amount of information given in the descriptions, and in use of jargon terms that embraced several concepts at once. It would be interesting to assess the extent to which each of these factors contributed to the overall difficulty of the task.

To verify the findings of this research with regard to cost consciousness and its impact on search performance and heuristic approach, it would be necessary to repeat the study on databases with online charges that are significant to experienced searchers. They did not perceive the costs for the ONTAP database as being significant compared to those they experienced in their work situation, although they did respond to the treatment. The question raised, then, is how they respond to a cost constraint when searching a costly database. Would expert searchers respond to the treatment by cutting costs in the same way that they did in this study?

The effect of anxiety on searchers' performance was inconclusive. It is a factor that deserves further analysis, particularly in looking at anxiety levels at different points in the online session. It would also be useful to try to partition anxiety and compare the effects of specific sources of anxiety on search performance.

Another area of interest is the ability of searchers to assess the 'goodness' of their answer sets and the likelihood of success with their search terms. Given that all the subjects in this study were relatively unfamiliar with the database, it would be of interest to determine success rates for prediction of the 'goodness' of search terms and evaluation of answer sets on more frequently used, more familiar databases. It would be expected that expert searchers searching on familiar databases would be better able to make an estimate of number of references in the database relevant to a question. If this were the case, then the question of interest would be whether the accuracy of the estimate affects the success of the search.

An area that deserves further exploration is that of search strategy selection. It would be useful to study the factors that go into selecting a search strategy for a particular question. Courses in online searching often give guidelines to selecting a search method for a question, but whether or not searchers apply the guidelines is not really known.

It would also be useful to compare the decision making process for novices and experts. One area for concern would be whether searchers tend to cling to certain strategies and apply them without regard to the nature of the query. This research indicated that searchers were more than willing to discuss their choice of strategy before beginning the search, giving their reasoning in

great detail. Therefore, study of the decision making involved would probably pose little difficulty in terms of methodology. There was also a tendency for searchers to want to discuss the strategy and modifications made to the search at its completion, and this would also provide valuable information about the search process.

The method of having searchers verbalize their thoughts while searching provides tremendous insight into the search process. Though not a readily quantifiable method, the results go a long way toward elucidating the strengths and weaknesses of searchers' problem solving. It is strongly recommended that the behaviour of novice searchers be studied more extensively, both novice intermediary searchers and professional end-users. It will provide valuable information for use in designing better interfaces and in development of expert systems.

CONCLUSIONS

The purpose of this research was to attempt to identify factors affecting the heuristic approach of searchers while interacting with an online system, and to relate that behaviour to search success in the form of enhanced recall. Several important findings emerged from this study.

Cost consciousness on the part of the searcher was not found to significantly influence heuristic approach, nor

did it influence recall attained. The treatment imposed on the subjects was found to be effective in eliciting feelings of cost consciousness, and these feelings translated into a more efficient search style. However, this alteration in search process did not significantly reduce search success. An important finding from this research is that heuristic approach, as defined and measured in this study, was not suppressed by cost consciousness, as has often been suggested in the literature of online searching.

There is evidence from this study that higher recall resulted from a high heuristic search approach. However, the association was weaker than expected, and was dependent upon the search question. A basic assumption about online searching is that the more interactive the searcher and the more heuristic the search approach, the better the search will be. The findings of this research indicate that this may be true for some search questions and not for others. It is important to note that a "quick-and-dirty" search approach yielded high recall for one of the two search questions in this study.

Measuring heuristicity proved to be problematic for this research for a number of reasons. Primarily, the difficulty lies in trying to quantify factors such as flexibility, adaptability, and gathering and utilization of feedback from the search system. However, it is clear that heuristicity is an important part of search behaviour and

more effort should be placed on finding a valid measure for this phenomenon.

The nature of the search question emerged over and over again in the analyses as having a main effect on the search process. Care should be taken in any research into online searching that generalizes beyond the specific search tasks being performed. Studying search behaviour for questions having different characteristics should prove highly illuminating, and would be essential in gathering data for developing an expert system intermediary.

It was apparent from this study that the method of verbal protocols represents a valuable tool for the study of online searching. Listening to the thoughts of experts and novices while interacting with the search system was illuminating. Surprising differences and similarities emerged when a large number of subjects searched the same search questions. The information obtained using this method would be valuable in designing online courses, in designing online systems, and in developing new research methodologies. Perhaps the greatest value of this methodology lies in its power to uncover deficiencies in problem solving approach that hamper searchers. Errors in strategy can be gleaned simply by viewing transaction logs; however, deficiencies in problem solving can only be identified by listening to thoughts verbalized during the problem solving process. It becomes quickly apparent when a searcher is losing sight of the final goal, becoming

fixed on a sub-goal, or choosing inappropriate goals. It is this process that is often overlooked in teaching online searching and in developing help systems for online intervention. Verbal protocols represent a powerful tool for future research in this field.

Finally, the need to look at the entire search process, rather than just the online session, was apparent. It was clear from this research that the search process was constrained by denying the searcher interaction with the requestor, and it appeared that some subjects were more affected by this than others. It is, therefore, important when conducting this kind of research to allow the subjects to express their feelings about the constraints of the experimental situation. It was clear from their verbalizations that most searchers considered the search task to be a first iteration of an ongoing problem solving exercise. To develop a true picture of online searching as a problem solving process, it is necessary to include the pre- and post-search interaction with the requestor, to take into account the situational variables of the searcher, and to allow for the iterative nature of the online search process.

APPENDIX 1

Forms and Questionnaires

To the Participant,

DESCRIPTION OF THE RESEARCH

This research project involves studying how people conduct online searches. It entails doing three searches on the ONTAP ERIC database of the DIALOG online search service. While conducting the searches, you will be asked to speak aloud your thoughts. You will be audiotaped throughout each search.

You will be asked to complete a background questionnaire that deals mostly with your past search experience. In addition, for each question to be searched, you will (1) work out your search strategy on the paper provided, (2) answer a brief questionnaire, (3) carry out the search at the terminal, and (4) complete a second brief questionnaire. You will be told when to begin each of these four tasks, and may take as much time as you need to complete them. Both the ERIC Thesaurus and a brief guide to searching DIALOG will be availabl to you at all times.

Before beginning the searching, you will do an exercise to practice thinking aloud.

The entire session should take approximately an hour. Any questions about this research should be directed to Jill Austin at (613) 728-5453.

Subject:

PARTICIPANT CONSENT FORM

I have read the description of the research and I agree to participate in this research project with the understanding that I will not be subjected to any physical discomfort, and with the further understanding that any information generated shall be reported in such a way that strict anonymity will be maintained for all subjects.

It is also understood that I will be free to withdraw from the experiment at any time without prejudice.

_____ _____

Date Signature

BACKGROUND QUESTIONNAIRE **Subject:** **Date:**

Current Occupation _____

Sex _____

Education:

Please list post-secondary degrees:

Degree	Subject Area	Date Completed

If you are currently registered in the MLIS program, complete the following, what term are you in? _____

Computer Experience:

How many computing courses have you taken?

Do you use a computer on a regular basis? ____ Yes ____ No

If yes,

How frequently do you use it? ____ times per month

How long have you been using it? ____ months, ____ years

What do you use it for?

_____ Coursework _____ Programming

_____ Games _____ Statistical Work

_____ Text Processing _____ Online Searching

_____ Other (list) _____

Library Experience:

Have you ever been employed in a library or information centre?

_____ Yes _____ No

If yes,

How long in total? _____ months, _____ years

Are you still employed in a library or information centre?

_____ Yes _____ No

If no, give last year you were so employed _____

List the activities you were primarily involved in:

How would you rate yourself in experience in conducting manual literature searches? Circle the appropriate number below.

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - - 7
No Experience — Very Experienced

Online Search Experience:

The following questions relate to your online experience.

Was it done _____ as part of coursework?

_____ as part of your job?

_____ for your own use?

Your last search was within the past

_____ 1 month? _____ 2 years?

_____ 6 months? _____ 5 years?

_____ 1 year?

How much searching, on average, do (did) you do?

_____ times per month, _____ times per year

How would you rate yourself in terms of online searching experience? Circle the most appropriate number below.

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - - 7
No Very
Experience Experienced

If you currently do online searching in your job, describe the charging polices used

What proportion of your searching experience has been on DIALOG?

_____ < 10% _____ 51-80%

_____ 10-50% _____ > 80%

Have you ever searched the ERIC database online?

_____ Yes _____ No

If yes, what proportion of your searching has been done on the ERIC database?

_____ < 10% _____ 51-80%

_____ 10-50% _____ > 80%

Have you ever used the ONTAP ERIC training database?

_____ Yes _____ No

Subject: Date: Task: 1 2 Question:

PRE-SEARCH QUESTIONNAIRE

Note: When a scale is given, circle the number that best describes your answer.

1. How familiar are you with the subject area of this question?

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Familiar — Very Familiar

2. To what extent do you feel that lack of interaction with the requestor is detrimental to your ability to search this question?

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Detrimental — Very Detrimental

Explain (if necessary)_____

3. Do you have some idea of the number of relevant references you expect to find? ____ Yes ____ No

If yes, about how many? _____

4. How confident are you that you have uncovered the best terms to search for?

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Confident — Very Confident

Subject: Date: Task: 1 2 Question:

POST-SEARCH QUESTIONNAIRE

Note: When a scale is given, circle the number that best describes your answer.

1. How confident are you that you have uncovered all the references relevant to this question?

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Very
Confident Confident

2. How satisfied are you with the content of your answer set?

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Very
Satisfied Satisfied

Explain (if necessary)_____

3. Do you feel your answer set is ___ Too large?

___ Too small? ___ About right?

4. Describe your level of anxiety while searching this question.

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Very
Anxious Anxious

Explain (if necessary)_____

5. To what extent were you concerned about cost while you were searching?

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Very
Concerned Concerned

6. To what extent do you feel you could have improved the search if you had been able to discuss the question with the requestor?

1- - - - 2- - - - 3- - - - 4- - - - 5- - - - 6- - - -7
Not Very
At All Much

Explain (if necessary)_____

Subject: **Task: Test** **Question: S09**

SEARCH QUESTION

Parapsychology. Find everything in the database relevant to this topic.

Subject: **Task:** 1 2 **Question:**

SEARCH QUESTION

Library service to the physically handicapped (not mentally or language handicapped). Find everything in the database that is relevant to this topic.

Subject: **Task:** 1 2 **Question:**

SEARCH QUESTION

White flight to the suburbs. Find everything in the database that is relevant to this topic.

Subject: Date:

SEARCH DATA

Question number: S09

Final set #: No. refs in final set:

Answer set #: No. refs in answer set:

No. common refs: Cost:

Recall: Precision:

Question number:

Final set #: No. refs in final set:

Answer set #: No. refs in answer set:

No. common refs: Cost:

Recall: Precision:

Question number:

Final set #: No. refs in final set:

Answer set #: No. refs in answer set:

No. common refs: Cost:

Recall: Precision:

Comments:

APPENDIX 2

Experimental Procedures

(Treatment)

INSTRUCTIONS TO PARTICIPANT

You will be carrying out three searches on the ONTAP ERIC database of the DIALOG online search service. This is a special training database consisting of one year (1975) of the ERIC database, containing approximately 35,000 records.

For each question to be searched, you will (1) work out your search strategy on the paper provided, (2) answer a brief questionnaire, (3) carry out the search at the terminal, and (4) complete a second brief questionnaire. You will be told when to begin each of these four tasks, and may take as much time as you need to complete them. Both the ERIC Thesaurus and a brief guide to searching DIALOG will be available to you at all times. Each search should be as thorough as possible, attempting to find all documents relevant to the question. You should try to keep the cost of each search to a minimum.

You will not have an opportunity to interact with the person requesting the search, which may make the search more difficult and frustrating for you: just do the best you can in interpreting the search request.

While you are interacting with the computer, conducting your search, you are asked to 'think out loud.' You are not to describe your actions, but verbalize your thoughts as they occur to you. This will be recorded on audiotape. I will be in the room while you are searching, but will not answer any questions about the search itself. If you forget to talk aloud during your search, I will prompt you to talk. Try not to be self-conscious and to conduct your search as you would if you were alone, with the difference being that you will speak your thoughts aloud instead of to yourself.

When the search is finished, **do not log off.** You should tell me you are finished, give me the number of your final answer set, enter the command COST and tell me the cost of the search, then await further instructions.

Before beginning the searching, we will do a practice test in thinking aloud.

Do you have any questions?

(Control)

INSTRUCTIONS TO PARTICIPANT

You will be carrying out three searches on the ONTAP ERIC database of the DIALOG online search service. This is a special training database consisting of one year (1975) of the ERIC database, containing approximately 35,000 records.

For each question to be searched, you will (1) work out your search strategy on the paper provided, (2) answer a brief questionnaire, (3) carry out the search at the terminal, and (4) complete a second brief questionnaire. You will be told when to begin each of these four tasks, and may take as much time as you need to complete them. Both the ERIC Thesaurus and a brief guide to searching DIALOG will be available to you at all times. Each search should be as thorough as possible, attempting to find all documents relevant to the question. Do not be concerned about your search time or the cost.

You will not have an opportunity to interact with the person requesting the search, which may make the search more difficult and frustrating for you: just do the best you can in interpreting the search request.

While you are interacting with the computer, conducting your search, you are asked to 'think out loud.' You are not to describe your actions, but verbalize your thoughts as they occur to you. This will be recorded on audiotape. I will be in the room while you are searching, but will not answer any questions about the search itself. If you forget to talk aloud during your search, I will prompt you to talk. Try not to be self-conscious and to conduct your search as you would if you were alone, with the difference being that you will speak your thoughts aloud instead of to yourself.

When the search is finished, do not log off. You should tell me you are finished and give me the number of your final answer set, then await further instructions.

Before beginning the searching, we will do a practice test in thinking aloud.

Do you have any questions?

TALK ALOUD PRACTICE QUESTION

Before you begin searching, I am going to give you a chance to practice thinking aloud while you do a problem.

I will give you two numbers to multiply together in your head. Do not describe what you are doing: instead, I want you to verbalize your thoughts, talking aloud as you work out the answer.

First problem: 12 times 16 Answer: 192

Second problem: 32 times 7 Answer: 224

Good. Now we can go ahead with the first search.

APPENDIX 3

Sample Complete Log

I'm going to start with the two broadest descriptors, I think, and see what (inaudible) words and sets are.

SS library(w)services/de and physically(w)handicap?/de

OK. Library services seems to be a good one, but physically handicapped...

Maybe it should be physical handicapped or disabled.

Though I'm not convinced that disabled is just physically handicapped.

I couldn't tell from the book.

SS disabled/de or physical(w)handicaps/de

Well, there's quite abit.

Physical handicaps is not bad.

Combine library services, which is S?, and S6, which was the original physically handicapped, and see what I get.

SS s3 and s6

Nothing? Of course, it was zero.

SS s3 and s8

(inaudible) disabled. There's nothing under disabled either. Hmm.

S s3 and s10

I'm going to try it without descriptor, the disabled part.

SS s3 and disabled

SS s3 and handicap?

Seems library services and handicapped is best.

I'll just look at a few and see what their terms are.

T s19/8/1-3

Disabilities, ahh.

SS s3 and disabilities/de

(inaudible) down to nine.

SS s21 and physical

Check the descriptors again.

T s23/8/all

SS s3 and physical(w)disabilities/de

Same three (?)

(checking thesaurus (?))

There doesn't seem to be very much when you use the sort of global term physical disabilities...

But without talking to the person who's giving the question, I don't know how much they want.

I could put in library services and list different types of physical disabilities, if that's what they want, but at this point I don't know.

So once again, I'd stop and tell them what I got and tell them we'd have to, they'd have to be more specific about what they want.

I suppose I could combine (inaudible) S21.

Let me just check. S21? S19? S12.

SS s27 or s21 or s19 or s12

I just combined the sets that I thought would have the most, and set 28, with 27 hits.

APPENDIX 4

Instructions to Coders

CODING SEARCH TRANSCRIPTS

You will be coding only the SELECT and SELECT STEPS commands (S and SS). In addition, you will only be coding S or SS commands that contain search terms (if they contain only sets, e.g., SS S5 OR S10, do not code them).

Use as your decision criterion the verbalizations of the subject and the search form filled out by the subject before the search session. Some judgement is required in using the search form information: the subjects worked out a pre-search formulation on the search form, but they also used the search form to jot down terms as they saw them or thought of them during the search. The terms added during the search may be less neatly written, will probably be lower on the sheet, and may not show coordination between terms.

ERRORS

When an error is already coded, LE (logical error) or CE (command error), do not code the command. The error code will appear in parentheses at the end of the command line on the coding form.

When trying to code a command that the subject has redone due to an error (sometimes several times), use as your criterion the motivation for the original command.

S CODES:

Search for multiple terms in a single command, mark any applicable codes, e.g., if there are 3 terms and 1 is an S1 and the others are S2, mark S1, S2 x 2. When different types of terms occur in the same command, mark above each term its type. If a term has been entered identically more than once, then code only the first instance.

S1 - search for term(s) from initial pre-search formulation. If the subject begins the search with a term from the query, that he/she has not written on the search form, code it as S1 - it may have been too obvious to write down.

S2 - altering a previously searched term in some way (adding or deleting adjacency operators, field restrictions, major designation, part of the term, or an affix). This does not include variations of a term given in the pre-search formulation, which should be coded as S1.

Ex: physically(w)(handicapped or disabled) = 2 terms
changed to
physically and (handicapped or disabled) = S2 x 3

S3 - searching for a term found by viewing citations

S4 - searching for a term just thought up in the process of searching (neither S1 nor S2 nor S3). If none of the other codes apply, use S4. If you suspect that the term was found from viewing citations (S3), but the subject did not verbalize it as such, use the S4 code. Use this code for terms that were not used in the initial formulation, e.g., parapsychology, but were tried later in the search.

S5 - experimenting to learn more about how the system works

APPENDIX 5

Descriptive Statistics

TABLE A1: Descriptive Statistics for Search Outcome and Heuristic Index Variables

	N	Mean	StdDev	Minimum	Maximum	Range
Question 2						
Recall	39	45.513	21.739	0.0	85.0	85.0
Precision	39	52.436	35.786	0.0	100.0	100.0
Total Cost	40	3.85	1.98	1.12	10.43	9.31
Recall*	39	41.362	15.413	0.0	67.21	67.21
Precision*	39	49.084	28.897	0.0	90.00	90.00
Heuristic Index	39	-.217	.609	-1.100	1.410	2.51
Question 3						
Recall	37	48.649	36.981	0.0	100.0	100.0
Precision	37	26.405	30.051	0.0	80.0	80.0
Total Cost	40	4.05	1.86	1.77	9.48	7.71
Recall*	37	43.033	29.040	0.0	90.0	90.0
Precision*	37	26.289	22.556	0.0	63.44	63.44
Heuristic Index	37	0.080	.774	-.96	2.230	3.19

* Arcsin square-root transformation.

TABLE A2: Descriptive Statistics for Search Process Variables

	N	Mean	StdDev	Minimum	Maximum	Range
Question 2						
Bool. ORs	40	4.275	5.765	0.0	34.0	34.0
Bool. ANDs	40	2.475	1.601	0.0	9.0	9.0
Commands	40	10.80	5.441	2.0	22.0	20.0
Orig.Terms	40	5.60	4.442	2.0	28.0	26.0
New Terms	40	2.80	2.997	0.0	15.0	15.0
Cites Viewed	40	7.30	5.662	0.0	20.0	20.0
Sets Browsed	40	1.65	.975	0.0	4.0	4.0
Cycles	40	1.70	.992	0.0	4.0	4.0
Complexity	39	11.333	10.922	2.0	51.0	49.0
Question 3						
Bool. ORs	40	4.200	3.560	0.0	15.0	15.0
Bool. ANDs	40	3.175	2.500	0.0	11.0	11.0
Commands	40	12.075	6.681	3.0	37.0	34.0
Orig.Terms	40	5.300	3.510	1.0	16.0	15.0
New Terms	40	2.500	2.774	0.0	12.0	12.0
Cites Viewed	40	10.875	11.182	0.0	52.0	52.0
Sets Browsed	40	2.350	1.272	0.0	6.0	6.0
Cycles	40	2.450	1.395	1.0	6.0	6.0
Complexity	37	7.541	5.684	1.0	20.0	19.0

TABLE A3: Descriptive Statistics for Other Search-Specific Variables (Question 2)

	N	Mean	StdDev	Minimum	Maximum	Range
Topic Familiarity (1=not familiar, 7=very familiar)	40	4.075	1.655	1.0	7.0	6.0
Lack of Requestor Interaction (1=not detrimental, 7=very detrimental)	40	3.600	1.374	1.0	6.0	5.0
Idea of Number of Relevant Citations (1=yes, 2=no)	40	1.725	.452	1.0	2.0	1.0
Estimated Number of Citations Expected	38	12.947	40.395	0.0	200.0	200.0
Using Best Search Terms (1=not confident, 7=very confident)	39	5.103	1.119	2.0	7.0	5.0
Obtained All Relevant Citations (1=not confident, 7=very confident)	40	4.225	1.672	1.0	7.0	6.0
Content of Answer Set (1=not satisfactory, 7=very satisfactory)	40	4.350	1.388	1.0	6.0	5.0
Size of Answer Set (0=unknown, 1=too large, 2=too small, 3=about right)	40	2.450	.749	0.0	3.0	3.0
Anxiety During Search (1=not anxious, 7=very anxious)	39	3.410	1.728	1.0	6.0	5.0
Cost Consciousness (1=not concerned, 7=very concerned)	40	2.575	1.796	1.0	6.0	5.0
More Interaction With Requestor Would Improve Search (1=not at all, 7=very much)	39	4.308	1.852	1.0	7.0	6.0

TABLE A4: Descriptive Statistics for Other Search-Specific Variables (Question 3)

	N	Mean	StdDev	Minimum	Maximum	Range
Topic Familiarity (1=not familiar, 7=very familiar)	40	2.575	1.448	1.0	6.0	5.0
Lack of Requestor Interaction (1=not detrimental, 7=very detrimental)	40	5.325	1.845	1.0	7.0	6.0
Idea of Number of Relevant Citations (1=yes, 2=no)	40	1.775	.423	1.0	2.0	1.0
Estimated Number of Citations Expected	39	2.718	8.249	0.0	45.0	45.0
Using Best Search Terms (1=not confident, 7=very confident)	40	3.250	1.645	1.0	7.0	6.0
Obtained All Relevant Citations (1=not confident, 7=very confident)	40	3.525	1.739	1.0	6.0	5.0
Content of Answer Set (1=not satisfactory, 7=very satisfactory)	40	3.650	1.626	1.0	7.0	6.0
Size of Answer Set (0=unknown, 1=too large, 2=too small, 3=about right)	40	2.200	.723	0.0	3.0	3.0
Anxiety During Search (1=not anxious, 7=very anxious)	40	3.900	1.766	1.0	7.0	6.0
Cost Consciousness (1=not concerned, 7=very concerned)	39	2.795	1.908	1.0	7.0	6.0
More Interaction With Requestor Would Improve Search (1=not at all, 7=very much)	39	5.821	1.636	1.0	7.0	6.0

APPENDIX 6

Additional Tables of Results

TABLE B1(a): Rank Order Correlation Matrix of Self-rated Variables*

	Familiarity With Search Topic	Lack of Requestor Interaction	Using Best Search Terms
Lack of Requestor Interaction	-.0007 (-.5054)	---	---
Using Best Search Terms	.5206 (.3644)	-.0193 (-.5076)	---
Found All Relevant Cites	.3167 (.2847)	-.0350 (-.2866)	.3564 (.3491)
Satisfaction With Answer Set	.0945 (.2141)	.0337 (-.0442)	.2921 (.1430)
Anxiety Level	-.1974 (.0673)	.1055 (.0367)	.0470 (-.2296)
Cost Consciousness	-.1342 (.0915)	.2387 (.1049)	-.0849 (-.1629)
Need For More Requestor Interaction	.0885 (-.4318)	.3528 (.7365)	-.1215 (-.4555)

* The first value given is for search question 2; the second value given, in parentheses below the first, is for search question 3.

TABLE B1(b): Rank Order Correlation Matrix of Self-rated Variables*

	Found All Relevant Cites	Satisfaction With Answer Set	Anxiety Level	Cost Consciousness
Satisfaction With Answer	.0377 (.7025)	---	---	---
Anxiety Level	-.0570 (-.2241)	-.1911 (-.1174)	---	---
Cost Consciousness	-.1449 (-.0989)	-.2374 (.0434)	.3831 (.3383)	---
Need For More Requestor Interaction	.0091 (-.3526)	-.3745 (-.0455)	.4095 (-.0022)	.1579 (.1007)

* The first value given is for search question 2; the second value given, in parentheses below the first, is for search question 3.

TABLE B2: Association Between Topic Familiarity and Search Process and Outcome Variables

	Familiarity with Search Topic	
	Question 2	**Question 3**
Search Outcome Variables		
Recall	-.0700	.1077
Precision	.0502	-.0031
Total cost	.0971	.0964
Heuristic Variable		
Heuristic Index	-.1946	.3353
Search Process Variables		
Boolean ORs	-.1839	.0266
Boolean ANDs	-.2198	.2486
Total commands	-.0996	.3327
Original terms	-.1103	-.0719
New terms	-.3678	.3687
Citations viewed	.0061	.1285
Sets browsed	-.0757	.1910
Cycles	.0075	.1686
Complexity	-.0999	.1918

TABLE B3: Association Between Anxiety and Search Process and Outcome Variables

	Self-rated Anxiety	
	Question 2	Question 3
Search Outcome Variables		
Recall	-.1161	-.0939
Precision	-.0828	-.0710
Total cost	.1313	.2969
Heuristic Variable		
Heuristic Index	.0305	.1784
Search Process Variables		
Boolean ORs	.0662	-.1246
Boolean ANDs	.1302	.4118
Total commands	-.0624	.0470
Original terms	.0542	-.0727
New terms	.1546	.0855
Citations viewed	.0000	.1318
Sets browsed	.0148	.2914
Cycles	-.0638	.2845
Complexity	.0090	-.0326

TABLE B4: Association Between Lack of Requestor Interaction, Search Process and Outcome, and Heuristic Index Variables

	Lack of Requestor Interaction	
	Question 2	Question 3
Search Outcome Variables		
Recall	-.0825	-.0268
Precision	-.4322	.2105
Total Cost	-.1228	-.2751
Heuristic Variable		
Heuristic Index	-.1343	-.4467
Search Process Variables		
Boolean ORs	.0246	-.3620
Boolean ANDs	-.1922	-.2433
Total commands	-.3047	-.3892
Original terms	.0849	-.1814
New terms	.0679	-.3654
Citations viewed	-.0616	-.2636
Sets browsed	-.2032	-.3684
Cycles	-.1809	-.3291
Complexity	.0340	-.2338

TABLE B5: Association Between Confidence in Search Terms, Search Process and Outcome, and Heuristic Index Variables

	Confidence in Search Terms	
	Question 2	Question 3
Search Outcome Variables		
Recall	.2588	-.1049
Precision	.0957	-.1017
Total Cost	.1090	.3596
Heuristic Variable		
Heuristic Index	.0717	.2773
Search Process Variables		
Boolean ORs	-.0925	.3531
Boolean ANDs	-.0177	.2585
Total commands	.0708	.3253
Original terms	-.1665	.2985
New terms	.0491	.1889
Citations viewed	.0444	.0471
Sets browsed	.0138	.3276
Cycles	.0600	.2694
Complexity	-.1190	.3333

TABLE B6: Association Between Confidence in Recall Attained, Search Process and Outcome, and Heuristic Index Variables

	All Relevant Citations Found	
	Question 2	Question 3
Search Outcome Variables		
Recall	.0747	.1177
Precision	-.2425	.1553
Total Cost	-.2739	-.1825
Heuristic Variable		
Heuristic Index	-.0981	-.0578
Search Process Variables		
Boolean ORs	.0649	-.1780
Boolean ANDs	-.1398	-.1360
Total commands	-.1433	-.0805
Original terms	.1631	-.0766
New terms	-.1593	-.0761
Citations viewed	.0015	.0157
Sets browsed	.0433	-.0124
Cycles	.0683	-.0275
Complexity	.0641	-.1846

TABLE B7: Association Between Satisfaction with Answer Set, Search Process and Outcome, and Heuristic Index Variables

	Satisfaction with Answer Set	
	Question 2	Question 3
Search Outcome Variables		
Recall	.2686	.3615
Precision	.1296	.3703
Total Cost	-.0347	-.1142
Heuristic Variable		
Heuristic Index	-.0857	-.0082
Search Process Variables		
Boolean ORs	-.0442	-.3830
Boolean ANDs	-.2513	-.0780
Total commands	-.1487	-.1144
Original terms	-.0666	-.1210
New terms	-.0864	-.1848
Citations viewed	.0608	.2084
Sets browsed	-.0623	.0962
Cycles	-.0361	.0955
Complexity	.0734	-.4680

TABLE B8: Association Between Need for More Requestor Interaction, Search Process and Outcome, and Heuristic Index Variables

	Need for More Requestor Interaction	
	Question 2	Question 3
Search Outcome Variables		
Recall	-.2922	-.0331
Precision	-.2221	.1877
Total Cost	.0836	-.0008
Heuristic Variable		
Heuristic Index	.0444	-.2397
Search Process Variables		
Boolean ORs	.0781	-.2362
Boolean ANDs	.0428	-.0849
Total commands	.0413	-.1490
Original terms	.0547	-.0508
New terms	.0756	-.1602
Citations viewed	-.1263	-.2165
Sets browsed	.0733	-.2884
Cycles	.0767	-.2063
Complexity	-.0115	-.1927

REFERENCES

Allen, B. (1990). Knowledge organization in an information retrieval task. Information Processing & Management, 26(4): 535-542.

Anders, V., & Jackson, K.M. (1988). Online vs. CD-ROM: the impact of CD-ROM databases upon a large online searching program. Online, 12(6): 24-32.

Baker, C.A., & Eason, K.D. (1981). An observational study of man-computer interaction using an online bibliographic information retrieval system. Online Review, 5(2): 121- 132.

Bates, M.J. (1979a). Information search tactics. Journal of the American Society for Information Science, 30: 205-214.

Bates, M.J. (1979b). Idea tactics. Journal of the American Society for Information Science, 30: 280-289.

Bates, M.J. (1981). Search techniques. Annual Review of Information Science and Technology, 16: 139-169.

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13(5): 407-424.

Bellardo, T. (1985). An investigation of online searcher traits and their relationship to search outcome. Journal of the American Society for Information Science, 36(4): 241-250.

Blackshaw, L., & Fischhoff, B. (1988). Decision making in online searching. Journal of the American Society for Information Science, 39(6): 369-389.

Blair, D.C. (1980). Searching biases in large interactive document retrieval systems. Journal of the American Society for Information Science, 31: 271-277.

Borgman, C.L. (1983). Performance effects of a user's mental model of an information retrieval system. Proceedings of the 46th American Society for Information Science, 20, 121-124.

Borgman, C.L. (1986a). Individual differences in the use of information retrieval systems: a pilot study. Proceedings of the 49th American Society for Information Science, 23: 20-31.

Brahmi, F.A. (1988). The effect of CD-ROM Medline on online end user and mediated searching. Medical Reference Services Quarterly, 7(4): 47-55.

Byrne, A. (1989). 'Or else they're lovely liars': online searchers in Australia. Online Review, 13(5): 395-405.

Dorner, D. (1983). Heuristics and cognition in complex systems. In R. Groner, M. Groner, & W.F. Bischof (Eds.), Methods of heuristics (pp. 89-107). Hillsdale, NJ: Lawrence Erlbaum.

Ericsson, K.A., & Simon, H.A. (1984). Protocol Analysis. Cambridge: MIT Press.

Fenichel, C.H. (1981). Online searching: measures that discriminate among users with different types of experiences. Journal of the American Society for Information Science, 32(1): 23-32.

Fenichel, C.H. (1980/81). The process of searching online bibliographic databases: a review of research. Library Research, 2: 107-127.

Fidel, R. (1983). Do user charges affect online searching behavior? Proceedings of the 46th America: Society for Information Science, 20: 132-134.

Fidel, R. (1984). Online searching styles: a case-study-based model of searching behavior. Journal of the American Society for Information Science, 35(4): 211-221.

Fidel, R. (1985). Individual variability in online searching behavior. American Society for Information Science Proceedings, pp. 69-72.

Fidel, R. (1985). Moves in online searching. Online Review, 9(1): 61-74.

Fidel, R. (1986). Towards expert systems for the selection of search keys. Journal of the American Society for Information Science, 37(1): 37-44.

Fidel, R. (1987). What is missing in research about online searching behavior? Canadian Journal of Information Science, 12(3/4): 54-61.

Fidel, R., & Soergel, D. (1983). Factors affecting online bibliographic retrieval: a conceptual framework for research. Journal of the American Society for Information Science, 34(3): 163-180.

Fidel, R. (1988). Factors affecting the selection of search keys. Proceedings of the 51st American Society for Information Science, Atlanta, Georgia, October 23-27, Vol. 25, pp. 76-79.

Groner, M., Groner, R., & Bischof, W.F. (1983). Approaches to heuristics: a historical review. In R. Groner, M. Groner, & W.F. Bischof (Eds.), Methods of heuristics (pp. 1-18). Hillsdale, NJ: Lawrence Erlbaum.

Halperin, M., & Renfro, P. (1988). Online vs. CD-ROM vs. onsite: high volume searching - considering the alternatives. Online, 12(6): 36-42.

Harris, M.A. (1986). Sequence analysis of moves in online searching. Canadian Journal of Information Science, 11(2): 35-56.

Harter, S.P. (1984). Online searching styles: an exploratory study. College & Research Libraries, 45(4): 249-258.

Harter, S.P. (1984). Scientific inquiry: a model for online searching. Journal of the American Society for Information Science, 35(2): 110-117.

Harter, S.P., & Jackson, S.M. (Summer, 1988). Optical disc systems in libraries: problems and issues. RQ, 27(4): 516-527.

Harter, S.P. (1990). Search term combinations and retrieval overlap: a proposed methodology and case study. Journal of the American Society for Information Science, 41(2): 132-146.

Howard, H. (1982). Measures that discriminate among online searchers with different training and experience. Online Review, 6(4): 315-327.

Ingwersen, P., Johansen, Th., & Timmermann, P. (1980). User-librarian negotiations and search procedures: a progress report. In O. Harbo, & L. Kajberg (Eds.), Theory and application of information research. Proceedings of the 2nd International Research Forum on Information Science, 3-6 August 1977, Royal School of Librarianship, Copenhagen, pp. 160-171.

Kantor, P.B. (1987). A model for the stopping behavior of users of online systems. Journal of the American Society for Information Science, 38(3): 211-214.

Kuhlthau, C.C., Turock, B.J., & Belvin, R.J. (1988). Facilitating information seeking through cognitive models of the search process. Proceedings of the 51st American Society for Information Science, Atlanta, Georgia, October 23-27, Vol. 25, pp. 70-75.

Logan, E.L., & Woelfl, N.N. (1986). Individual differences in online searching behavior of novice searchers. Proceedings of the 49th American Society for Information Science, Vol.23, pp. 163-165.

Logan, E.L. (1988). Learning modes and online behavior of novice searchers. Proceedings of the National Online Meeting, New York, May 10-12, pp. 205-211.

Logan, E. (1990). Cognitive styles and online behavior of novice searchers. Information Processing & Management, 26(4): 503-510.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 40(1): 54-66.

Marshall, D.B. (1980). To improve searching, check search results. Online, 4(3): 32-47.

Marshall, J.G. (1990). Diffusion of innovation theory and end-user searching. Library and Information Science Research, 12: 55-69.

Meadow, C.T. (1986). Problems of information science research - an opinion paper. Canadian Journal of Information Science, 11: 18-23.

Morris, A., Tseng, G.M., & Walton, K.P. (1989). MOSS: a prototype expert system for modifying online search strategies. Proceedings of the 13th International Online Information Meeting, London, 12-14 December 1989, pp. 415-434.

Newell, A., & Simon, H.A. (1972). Human Problem Solving. Prentice- Hall, Englewood Cliffs, NJ, 920p.

Oldroyd, B.K. (1984). Study of strategies used in online searching. 5: differences between the experienced and the inexperienced searcher. Online Review, 8(3): 233-244.

Oldroyd, B.K., & Citroen, C.L. (1977). Study of strategies used in on-line searching. OnLine Review, 1(4): 295-310.

Penniman, W.D. (1975). A stochastic process analysis of on-line user behavior. Proceedings of the 38th American Society for Information Science, Vol.12, pp. 147-148.

Penniman, W.D., & Dominick, W.D. (1980). Monitoring and evaluation of on-line information system usage. Information Processing & Management, 16: 17-35.

Reed, S.K. (1982). Cognition. Brooks/Cole, Monterey, CA.

Rice, R.E., & Borgman, C.L. (1983). The use of computer-monitored data in information science and communication research. Journal of the American Society for Information Science, 34(4): 247-256.

Rowley, J.E., & Butcher, D.R. (1989). The Searcher/Information Interface Project 2: manual and online searching - pilot study. Journal of Information Science, 15: 109-114.

Salomon, K., & Burgess, C. (1984). Patron presence during the online search: attitudes of university librarians. Online Review, 8(6): 549-558.

Saracevic, T., Kantor, P., Chamis, A.Y., & Trivison, D. A (1988). A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science, 39(3): 161-176.

Saracevic, T., & Kantor, P. (1988a). A study of information seeking and retrieving. II. Users, questions, and effectiveness. Journal of the American Society for Information Science, 39(3): 177-196.

Saracevic, T., & Kantor, P. (1988b). A study of information seeking and retrieving. III. Searchers, searches, and overlap. Journal of the American Society for Information Science, 39(3): 197-216.

Sewell, W., & Teitelbaum, S. (1986). Observations of end-user online searching behavior over eleven years. Journal of the American Society for Information Science, 37(4): 234-245.

Smith, L.C. (1976). Artificial intelligence in information retrieval systems. Information Processing & Management, 12: 189-222.

Streufert, S., & Streufert, S.C. (1978). Behavior in the complex environment. Washington, D.C.: V.H. Winston.

Tenopir, C. (1988). Search strategies for full text databases. Proceedings of the 51st American Society for Information Science, Atlanta, Georgia, October 23-27, Vol. 25, pp. 80-86.

Tolle, J.E. (1983). Understanding patrons use of online catalogs: transaction log analysis of the search method. Proceedings of the 46th American Society for Information Science, Vol.20, pp. 167-171.

Trivison, D., Chamis, A.Y., Saracevic, T., & Kantor, P. (1986). Effectiveness and efficiency of searchers in online searching: Preliminary results from a study of information seeking and retrieving. Proceedings of the 49th American Society for Information Science, Vol.23, pp. 341-349.

Vigil, P.J. (1983). The psychology of online searching. Journal of the American Society for Information Science, 34(4): 281-287.

Woelfl, N.N. (August, 1984). Individual differences in online search behavior: the effect of learning styles and cognitive abilities on process outcome. Doctoral dissertation, School of Information and Library Science, Case Western Reserve University.