RELEVANCE THRESHOLDS: A CONJUNCTIVE / DISJUNCTIVE MODEL OF END-USER COGNITION AS AN EVALUATIVE PROCESS

Howard F. Greisdorf, B.S., M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2000

APPROVED:

Brian O'Connor, Major Professor
Linda Schamber, Co-Major Professor
Amanda Spink, Committee Member
Jon Young, Committee Member
Philip M. Turner, Dean of the School of Library and Information Sciences
C. Neal Tate, Dean of the Robert B. Toulouse School of Graduate Studies

Greisdorf, Howard F., Relevance Thresholds: A Conjunctive/Disjunctive Model of End-User Cognition as an Evaluative Process. Doctor of Philosophy (Information Science), December 2000, 166 pp., 12 tables, 15 illustrations, references, 157 titles.

This investigation identifies end-user cognitive heuristics that facilitate judgment and evaluation during information retrieval (IR) system interactions. The study extends previous research surrounding relevance as a key construct for representing the value end-users ascribe to items retrieved from IR systems and the perceived effectiveness of such systems. The Lens Model of user cognition serves as the foundation for design and interpretation of the study; earlier research in problem solving, decision making, and attitude formation also contribute to the model and analysis.

A self reporting instrument collected evaluative responses from 32 end-users related to 1432 retrieved items in relation to five characteristics of each item: topical, pertinence, utility, systematic, and motivational levels of relevance. The nominal nature of the data collected led to non-parametric statistical analyses that indicated that end-user evaluation of retrieved items to resolve an information problem at hand is most likely a multi-stage process. That process appears to be a cognitive progression from topic to meaning (pertinence) to functionality (use).

Each step in end-user evaluative processing engages a cognitive hierarchy of heuristics that includes consideration (of appropriate cues), differentiation (the positive or negative aspects of those cues considered), and aggregation (the combination of differentiated cue aspects needed to render an evaluative label of the item in relation to the information problem at hand). While individuals may differ in their judgments and

ACKNOWLEDGEMENTS

The following study is the result of the ongoing guidance, support and encouragement provided by a group of people that unwittingly allowed me to recognize a scholarly passion that I have envisioned for many years, but only surfaced and gained momentum in the recent past.

To my mentor, Dr. Amanda Spink, I offer both praise and gratitude for instilling in me the need for scholarly aggressiveness, fortitude and understanding necessary to compete in an arena filled with the academic accomplishments and sometimes contrary ideas of others. As a champion of my abilities, she never wavered in challenging my thinking, my reasoning and my way of expressing ideas in order to get the best results out of my work. Her personal interest and constant encouragement provided the focus necessary for this current investigation.

I enthusiastically thank Dr. Brian O'Connor, as my committee co-chair, for acting in the capacity of wise counsel, ego-builder, scholarly logician, editor, and all-around advocate for my ideas. His continued confidence and inspiration added significant impetus to the timely completion of this work. Also, I extend my appreciation to Dr. Linda Schamber, committee co-chair, for her shared enthusiasm surrounding this investigation and her assistance in providing the academic grounding I required to begin this endeavor. I also thank Dr. Jon Young for his service on my committee in order to offer the interdisciplinary perspectives needed to generate the synthesis developed out of the nature of this study.

The underlying ideas leading to this investigation emanate from a conversation with Dr. Tefko Saracevic in the spring of 1997. At that time he said to me, “Looking for a single consistent individual human characteristic that explains the nature, manifestations and effects of relevance is like searching for the ‘Holy Grail’. I don’t think there is such a thing.” I’m not sure that I agree, so the challenge is in the search and this study represents my first step on that journey.

To all of those authors who have provided an underpinning for my approach to this investigation, my interpretation of the results and the succeeding discussion, I thank profusely with only one disclaimer. At all junctures I have tried to avoid misinterpretations or misrepresentations of the expressed work of others and apologize if any such errors have been implied by the statements made in this work.

Lastly, but certainly most important, I would like to thank my wife, Carole, for her continued support, encouragement, patience and endurance during this long and arduous process. Her constant interest, questions and suggestions acted as the pillars of strength I needed to bring this work to completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PROBLEM IDENTIFICATION

Introduction

Research involving the concept of relevance in the field of information science has produced a variety of significant findings that have caused an obvious bifurcation within the discipline as a whole. In one arena the focus has been on relevance as an evaluative binary measure of information retrieval effectiveness, while in the other, the focus has been on relevance as a cognitive manifestation of the information seeker encompassing a variety of aspects that extend beyond binary evaluation.

The goal of this research is to explore the nature and manifestations of relevance complexity as a cognitive evaluation process. The information seeker exhibits this complexity while attempting to resolve an information problem at hand by querying, searching, retrieving and evaluating items returned by an information retrieval system. This study focuses specifically on the user's subjective evaluation process.

In the context of this work, relevance complexity is considered to be a combination of cognitive components that information seekers utilize, not only to model their problem definitions as query formulations, but also to make decisions about retrieved information in order to move toward resolution of those information problems.

Research Objective

The objective of this study is to advance the research front encompassing the nature of user relevance judgments as evaluative interstices between the information retrieval process and the succeeding actions that a user may or may not take subsequent to such evaluations. This investigation assesses quantitatively a qualitative process in order to describe the conjunctive/disjunctive nature of relevance judging behavior as an evaluative tool in the realm of information science.

Robertson, Maron and Cooper (1982) identified three focused approaches to relevance as a relationship between a document and a topic, a relationship between a document and a search query and a relationship between a document and a person. Traditional approaches to theoretical perspectives surrounding the concept of relevance in information science, however, have generally suffered from a philosophical split that has divided these approaches into system-centered and user-centered research agendas. System-centered agendas have focused mainly on the first two relationships identified by Robertson et al., while the user-centered agendas have mainly addressed the relationship of user to document. This division has been well documented in the literature by Dervin & Nilan (1986), Ingwersen (1987), Schamber (1994), Froelich (1994) and Saracevic (1995, 1999) and represents a stimulus for this current study.

These camps have generally treated relevance as an evaluative process with two distinct frameworks for investigation that could be described as:

1) A system-centered physical approach focused predominantly on information retrieval in the context of system algorithmic constraints based on objective binary processing, i.e. parsing and matching query terms to document terms involving weighted and ranked output;

2) A user-centered cognitive approach focused primarily on information seeking, searching, retrieving and evaluating as a complex subjective human behavioral process.

This study posits that a binary definition of relevance from a systems design perspective and the conjunctive/disjunctive nature of human choice behavior act as variables in common that can bridge that gap between system-centered and user-centered research agendas. In order for these two divergent research agendas to effectively bring user and system together, a framework needs to be developed to bridge the gap created when a user disengages from the relationship of query to document and engages with the relationship between document and information need. This investigation explores two concepts that could possibly provide avenues for creating common denominators. These concepts postulate that:

1) Relevance is a subjective evaluative process consisting of a range of possible values; and
2) The conjunctive-disjunctive dichotomous nature of this evaluative process could objectively define that range of values.

Problem Statement

Subjective user cognition consistently acts in the capacity of evaluative interstices during human information seeking, searching and retrieving processes which has been succinctly posited by Marchionini (1995, p. 50) in his model of the information seeking process. Research in the realm of information science, however, has generally sought to identify and characterize only single variable concepts as major contributors to those evaluative events.

Prior research surrounding the concept of relevance in information science has posed numerous frameworks for describing and explaining the nature and manifestations of this concept to include considerations of criteria, value, utility, pertinence, topicality, goal orientation, situation, cognition, novelty, motivation and a variety of similar single variable contexts. Each of these aspects, attributes or dimensions of relevance has contributed to an overarching framework that may point to relevance as a concept that may incorporate all of these concepts for modeling human behavior in situations that require information evaluation in relation to an information-related problem at hand. Relevance is a complex human evaluative process that requires analysis that goes beyond single measure variables that attempt to define relevance as a whole.

Relevance judgment is an evaluative response by users that engenders a variety of criteria subject to the user's base of knowledge and the situation at hand. This study poses through the following research questions that there may be regions of relevance whose complexity may not be contingent on only the nature or number of criteria used in the evaluation process, but also on the positive versus negative aspects of the manifestations of relevance utilized for making those evaluative judgments.

Only a handful of studies have attempted to investigate this concept in relation to relevance as an evaluative process. The groundwork may have been laid in the early stages of growth in information retrieval systems when Van Rijsbergen (1975) identified that information retrieval not only consists of information extraction but in adequately defining how to use it to decide relevance. Wilson (1978) recognized that an epistemology surrounding individual belief systems could not be sorted out into two neat categories consisting of the true and the false. His insistence for a full range of possibilities between these extremes was important but not fully explained. At the same time Cooper (1978) spoke to the issue of positive and negative utility, yet concluded with a suggestion for abandoning attempts at more effective evaluation, as he moved ahead with an agenda focused on ranked models for retrieval.

In his discussion of meaning, implication and relevance, Overton (1990) identifies relevance as a bridge between meaning and implication in the sense of conditional logic (if p, then q). This position by Overton further supports relevance as, at least, a two stage process that operates cognitively as a structured whole. Most studies to date in information science have only looked at the structured whole (judgments as evaluative responses) without examining the separate stages.

More recently concepts of conjunction and disjunction have re-emerged as possible approaches for enhanced retrieval effectiveness. Blair (1996) in revisiting the results of an earlier investigation during the mid 1980's posits a system of evaluation based on complete conjunctive normal form for delineating subsets of document collections. His conclusion, however, is that such a methodology would be too cumbersome for large-scale systems. Blair, however, was only considering the conjunctive/disjunctive nature of document terms and not the same heuristic in terms of user relevance judgments.

Meghini, Sebastiani and Straccia (1998) in their work surrounding the logic of multimedia information retrieval postulate that the semantics of a concept can be understood as positive and negative extensions of the concept. This approach has been one of the few that articulates "that to arrive at a successful logical model of IR every effort should be made in order to wire as much relevance as possible into the implication connective. This means designing a calculus for (non-probabilistic) conditional reasoning where the factors that influence relevance, as perceived by the user, are taken into account."

Recently Mizzaro (1998) has attempted to posit relevance as a concept that resides in a four dimensional space, yet readily admits to the complexity of associations composing his fourth dimension of topic, task and context. He identifies this realm as the locus for the evaluation of results (relevance), but is unable to provide a clear understanding of how the combination of topic, task and context is cognitively interpreted and manipulated by users to reach an evaluative decision (relevance judgment). Mizzaro concludes by indicating that it is mandatory to proceed in an experimental way by confronting different kinds of relevance.

This study acknowledges and accepts that challenge by distinguishing between levels and regions of relevance in the context of human evaluative behavior and how these different kinds of relevance relate to human evaluations of information.

Importance of the Problem

In the field of information science, the concept of relevance has remained a bulwark for research that attempts to describe and explain the effectiveness of human interactions with IR systems. Historically, the emergence of this concept was for the purpose of providing units of measurement that could adequately assess IR system performance (Kent et. al., 1955). With time and continuing research focused on both system performance and human behavior in information environments, the conceptual framework surrounding relevance as an effective approach to measurement has been changing.

Schamber (1994), while providing a comprehensive compilation of diverse approaches to models and frameworks of relevance, also recognized that the discipline of information science has failed to reach any consensus on the issues of human behavior and measurement as they relate to an overall theory of relevance. The most crucial area for continuing investigation as posited by Schamber is being able to understand what users want when they interact with an IR system.

The importance of that position for moving the relevance research front was further reinforced by Froehlich (1994). He acknowledged that the process of relevance judging entails a system of relevances that relates the user and his/her need; and, what needs to be clarified are the interpretive schemes that users bring to an IR system.

Addressing the nature of this important problem, Saracevic (1996) clearly demonstrated that the effectiveness of IR depends on the interaction of various relevance manifestations organized into a system of relevances, and that ongoing relevance research in information science should be focused on that subject.

Research Questions

It is the intent of this study to address these important issues with an approach designed to examine:

1. How users cognitively orient themselves to items retrieved from an IR system to resolve an information need;
2. The complex nature of those various orientations;
3. How those complex orientations (levels of relevance) contribute to evaluations of importance (degrees of relevance) related to items retrieved and viewed by the user;
4. How the resulting constructs could lead to an enhanced theory of relevance in information science.

The specific questions addressed in this study are based on empirical results from prior work that establishes an underlying framework supporting the following givens:

1. End-users employ a wide range of categories to make decisions regarding the importance of information items retrieved from an IR system;
2. End-users are able to classify items retrieved from an IR system into categories (degrees of relevance) that encompass a range of values that extend beyond strict adherence to dichotomous evaluative judgments of relevance defined generally in the literature as relevant versus not relevant;
3. Human behavior associated with evaluative judgment is, for the most part, subjective, inconsistent and irreducible to single level variables.

If end-users are given an experimentally predefined set of relevance manifestations (levels of relevance) for assessing the relevance engendered by their evaluations of item importance during interactions with an IR system, do the end-users implement distinct consistent evaluative heuristics for determining regions of relevance?

Significance of the Research

Each of these research frameworks in information science could benefit from the results obtained in this investigation:

1) The relevance framework
2) The cognitive framework
3) The information seeking framework
4) The human information behavior framework

The Relevance Framework

While seeking to identify the key variables that influence the range of relevance judgments that users make in evaluating items retrieved from an IR system, most researchers have failed to recognize that each variable identified exists as a dichotomous evaluative attribute of relevance in both the positive and negative sense. In revisiting his earlier work, Blair (1996) is one of the few researchers to recognize this conjunctive-disjunctive relationship as a possible bridge between user and system at the document level. Others more recently have taken similar approaches in developing probabilistic approaches utilizing positive-negative manifestations of uncertainty to model logical solutions for more effective retrieval of information (Crestini, Lalmas & Van Rijsbergen, 1998). This binary positive versus negative nature of evaluative variables could be a key for connecting users more effectively to IR systems.

The Cognitive Framework

Although cognitive approaches to studies of relevance have flourished in the field of information science, most have focused on concepts that include knowledge states, learning styles, mental models, and problem solving as addressed effectively by Belkin (1978), Ingwersen (1982), Allen (1991), Harter (1992), and Mizzaro (1996). However, issues surrounding methods for describing and analyzing these cognitive states has remained relatively unexplored. The value of this investigation lies in its interdisciplinary application to concepts encompassed by the positive-negative conjunctive-disjunctive aspects of cognition in relation to relevance evaluative processes both in information science and in other disciplines concerned with human cognitive processes .

The Information Seeking Framework

The recent work of Bateman (1998) provides an extensive summary of information seeking, searching and retrieval models that engender relevance evaluative processes. While each of these models recognizes the cognitive nature of user judgments in the seek, search and retrieval process none has been able to identify how the user utilizes his/her discriminatory abilities to differentiate and integrate positive versus negative manifestations of relevance into their evaluative decisions of relevance. This study attempts to provide insights into that human cognitive process within an information seeking, searching and retrieval context.

The Human Information Behavior Framework

To establish and maintain efficient and effective heuristics for seeking, searching and retrieving information, humans coordinate a number of activities both mental and physical (Spink & Greisdorf, 1999). These coordinated elements, while including a user's cognitive state, knowledge state and problem state (Ingwersen, 1996), also include an evaluative state at key junctures of this coordination process. This study provides a capsulated analysis of one of these key junctures known in the field of information science as relevance judging. In a broader context, however, this investigation may add significantly to the total structure of human information coordinating behavior as a process approach to resolving an information need.

Assumptions and Terminology

Cognition

There remains a general consensus that the field of information science is divided into system-centered and user-centered research agendas. Similar classifications for this bifurcation have been represented as the objective versus subjective approach (Belkin, 1978) and the physical paradigm versus the cognitive paradigm (Ellis, 1992). This study is based on the assumption that human evaluative behavior exists within a cognitive domain consisting of a variety of dimensions that can provide a range of discriminating attributes that individuals are able to integrate into an evaluative response called relevance. This approach to the research treats relevance as an evaluative response not as a cognitive domain. Evaluation (judging the value, importance or worth of an item of retrieved information) is the cognitive domain. Relevance is a class or type of evaluative response stemming from aggregated judgments within that domain. Although the struggle in information science has been to resolve inconsistencies, ambiguities, and ambivalence in analyzing human information behavior surrounding relevance judging, the behavioral issue may not be relevance, but human evaluation behavior.

Measurement of retrieval success

The measurement of retrieval success has struggled with the fuzzy nature of relevance identified early by Rees (1966) as a confounding process of criteria, measures, measuring units, measuring instruments and methodology. Inconsistencies have abounded over the last 30 years with no real consensus on how users and systems can efficiently be brought together to assess information retrieval effectiveness for both the system and the user. This investigation posits that if relevance is used as such a measure of information retrieval effectiveness, then the nature of human evaluative behavior as an underlying framework needs to be included in any movement toward resolving this ongoing debate.

Degrees and Regions of relevance

The IR systems development community, for the most part, abides by a strict adherence to binary relevance logically consistent with the arguments posed by Cooper (1971), which indicate that relevance is an issue of “yes” or “no” without any intervening alternatives or gradations. While weighted and ranked outputs have attempted to connect users with their information needs more effectively, those outputs remain connected to query terms which may or may not adequately reflect user needs. Ongoing research from user perspectives, however, has built a strong argument in favor of a range of values or degrees of relevance that represents inferences related to a user's information need and items retrieved from an IR system to aid in resolving that need. This study assumes that a range of relevance values extends beyond strict binary relevance and represents aggregated evaluative responses based on the types or levels of relevance used in the evaluation process. These aggregated evaluative responses are defined in this study as regions of relevance. This distinction between degrees of relevance and regions of relevance is illustrated in Figure 1.1.

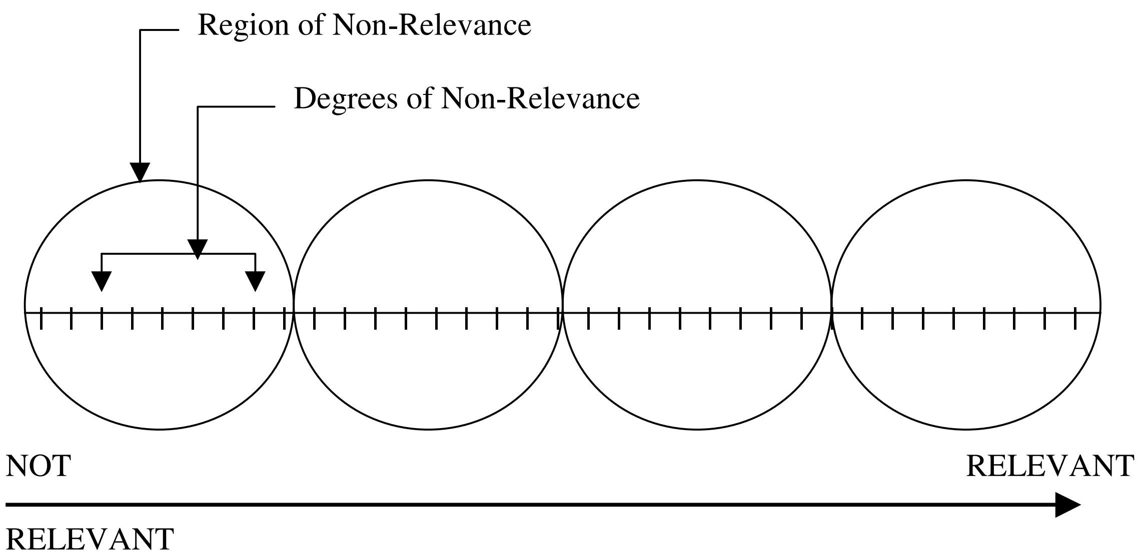

Figure 1.1 Distinction between degrees and regions of relevance

Levels of relevance

This research distinguishes between criteria, dimensions or factors of relevance (Schamber, 1991,1994; Barry, 1993; Wang & Soergel, 1993; Park, 1997; Barry & Schamber, 1998) and levels or manifestations of relevance (Saracevic 1996, 1999; Cosijn & Ingwersen, in press). This distinction posits that although users may have a vast array of judgment criteria at hand to make evaluative judgments, those criteria represent the personal biases and beliefs that influence their unique subjective behavior in environments and involving tasks where an information need requires resolution. On the other hand, levels of relevance represent aggregated inferences that are characteristic of items or objects of information as a whole. For example, a user may choose to use currency as a criterion for making a relevance judgment. Once retrieved, the relevance of the item may be judged on a next criterion. These singular or multiple 'finer grained' criteria, however, combine or aggregate to represent a judgment that engenders a next level of relevance represented by the criteria utilized for the retrieval process. These levels may include (but are not limited to) whether the item is on topic, how informative it is, how useful it is, whether it is in the right form or format, or whether it is a cause for further action in helping to resolve the information problem at hand.

Decision based dichotomies

Another underlying assumption in this study is that relevance can not be used as a binary, dichotomous, relevant versus not relevant variable in studies of user evaluative responses. Evaluative responses related to information retrieval have been shown to encompass a range of values that remove the concept of relevance from a strict dichotomous relationship of user to information need.

There exists, however, an interesting adjunct to this assumption. Humans do use dichotomous judgments in order to render evaluative responses. These cognitive dichotomies act as a differentiating and integrating function (Kelly, 1969). A difficulty in information science is that experimental methodologies have treated degrees of relevance as the judgments when they were only the evaluative response and not the judgments at all. The actual judgments surround the levels of relevance chosen by users to generate their evaluative responses. This study poses that it is the dichotomous nature of those levels that form the conjunctive and disjunctive combinations that generate those evaluative responses as regions of relevance.

Scope and Limitations

The scope of this investigation is limited to end-user evaluative behavior as exhibited following interaction with an IR system for the purpose of resolving an information need or problem. In the overall information seeking process this aspect of human behavior has generally been the focus of attention for determining IR system design effectiveness and therefore is a critical component of the research agenda in information science.

While some researchers have approached the process of information seeking as a key element in defining the nature of human information behavior (Wilson, 1999) this study recognizes the limitations imposed by only looking at one evaluative interstice in that total process. However, the value of this narrowly focused approach is to identify how the concept of binary or dichotomous relationships are integrated into user cognitive processes encompassing theoretical frameworks involving relevance as a unit of measurement in studies of IR system effectiveness and user interactions with such systems.

From a research design perspective, several issues are recognized as moderating variables that could influence the results obtained. These issues include the use of a convenience sample, the nature of the problems searched (domain knowledge of the user), number of searches (both single and successive), databases utilized (multiple), the number of items judged by each user and the type of measurement scales utilized. While these variables may place limitations on the results obtained, they also serve as real life interventions that effect the human evaluative process in relation to both information needs and IR system effectiveness.

Structure of the Research

This investigation was conducted by providing a self-reporting instrument (Appendix A) to 32 end-user graduate students for recording inferences of relevance as evaluative responses to items retrieved from IR systems to resolve their own information problems. A total of 1432 retrieved items constituted the corpus for analyzing end-user judgments and evaluations of relevance.

Summary

Information seeking, searching and retrieval activities by humans needing to fill a gap in their base of knowledge or make sense out of their current information based circumstances confront new information as an evaluation process. The judgments made as part of that process exist as manifestations or levels of importance (relevance) which users mix and match to help fill their knowledge gaps or make sense of their current state of affairs (situation). This study investigates both the positive and negative aspects of some specific levels of relevance considered to be representative of key dimensions of evaluative decision making when users interact with an IR system to retrieve or discover needed information. In addition, this study explores how users manipulate these positive and negative levels of relevance to create meaningful conjunctive and disjunctive combinations of these various levels of relevance. These combinations represent a range of importance (region of relevance) that corresponds to their current evaluative state resulting from their interaction with an IR system.

Based on the above evidence from the supporting literature, this investigation poses a model of relevance that treats user cognition as a multi-stage or multi-step process that aids in defining relevance complexity during user interactions with an IR system. This complexity emerges not at the early stages of information problem definition and query formulation, but at the point that search and retrieval results are examined by the user.

CHAPTER 2

HISTORICAL REVIEW AND CONTRIBUTING RESEARCH

Introduction

While a strict definition of relevance has remained elusive in the realm of information science, others have come to terms with the abstract nature of its semantic variability by introducing defining characteristics that appear to hold true across disciplinary approaches to this subject. These defining characteristics fall into two distinct categories that can be described as (1) what relevance is, and (2) what relevance is not.

Schutz (1970) speaks of relevance in terms of accomplishing the ‘diexodos’, choosing the alternatives in a situation of doubt, subjecting them to choice, each having its own weight, and giving assent by making a decision. It is not the questioning or doubting that leads to a decision, it is the assessment, based on convictions of certainty ranging from blind belief to opinion to indifference, of the problematic alternatives that produces a decision.

In his discussion of Popper’s philosophy, Settle (1976) identifies what he calls two obvious distinctions or substitutes for ‘certainty’ that he defines as the psychological and the epistemological. The psychological substitute for certainty is ‘belief’ while the epistemological substitute is ‘probability.’ Popper’s objectivist view of probability has been a mainstay in the IR system design arena as confirmed by Van Rijsbergen (1996), however, there may be a way of looking at user's actual beliefs that can deal satisfactorily with problems of knowledge, problems of uncertainty, and problems of relevance.

Vygotsky (1978) asserts that mere description does not reveal the actual dynamic relations that underlie phenomena. Descriptions of immediate experience in many cases do not provide us with an understanding of the real links between the external stimuli and the internal responses that underlie those higher forms of behavior classified by introspective descriptions. While relevance judgments may be representative of such introspective descriptions the underlying internal responses leading to those judgments require further investigation and understanding.

If relevance always comes from a pre-orientation within a background as posited by Winograd and Flores (1987), then investigations through experimentation need to explore the nature of these pre-orientations and backgrounds from which judgments of relevance emerge as evaluative responses to items retrieved from an IR system.

From a human communications approach, Sperber and Wilson's (1986) principle of relevance accounts for the interaction of linguistic meaning with interpretive utterance. In this investigation cognitive meaning derived from user interaction with an IR system could lead to interpretive utterances that engender underlying contextual factors of meaning. Sperber and Wilson further identify one of the key issues in the cognitive processing of information as the optimal allocation of central processing responses to meet the needs of the cognitive task. This optimization, created as a notion of contextual effects, is a necessary precursor for a condition of relevance. Such thinking is consistent with the methodology and analysis at the foundation of this study.

Within the realm of information science, relevance has been widely accepted as "the user's decision to accept or reject information retrieved from an information system (Schamber, 1994). The underlying assumptions inherent in such a definition further speak to what relevance represents as a concept. Relevance involves choice, it involves judgment and it culminates in a decision. This current study includes theoretical perspectives that bring those issues to bear on the nature and manifestations of relevance in the context of human evaluative behavior associated with interactions with IR systems.

Significant Conceptual Frameworks

Relevance

Operationalization of relevance as a framework for investigative study in information science has been confounded by the variety of semantic-laden descriptions utilized as relevance defining characteristics. This does not imply incorrect or misleading results from prior experimentation, but it does suggest caution for discriminating, differentiating and developing an orientation that more accurately identifies these semantic variations when approaching an investigative framework involving relevance related variables. The cause for such differentiations of meaning for the relevance construct stem from definitions in the literature that describe relevance as:

1) A psychological predicate which describes acceptance or rejection of a relation (Taube, 1965);
2) A criterion measure for quantifying the effectiveness of document retrieval systems (Rees, 1966);
3) A subjective notion (Van Rijsbergen, 1979);

4) A multidimensional cognitive concept (Schamber, Eisenberg & Nilan, 1990);
5) A complex systematic measurable concept (Schamber, Eisenberg & Nilan, 1990);
6) The fundamental criterion for evaluating the effectiveness of information retrieval (IR) and use (Schamber, 1994);
7) A built-in mechanism that comes along with cognition (Saracevic, 1996).

While some might argue that the above definitions are similar in context, an obvious inconsistency is evident. If relevance is a subjective cognitive notion initiated mentally for the purpose of evaluation, then it should be considered as a cognitive evaluative process, not a standard or criterion measure.

The dilemma associated with a clear consensus on the definition of relevance emerges from a variety of studies. These studies have treated terms such as measurement, judgment and evaluation as synonymous, when they could be viewed and treated as discrete concepts associated with the information seeking, searching and retrieval process. Constructs of relevance that have merged these discrete concepts into single variable approaches to experimentation may have contributed to confounding analyses of the results obtained. To clarify this viewpoint, the following additional terminology is presented:

Cognition

1) Any processing of information mediated by a system of categories of concepts (De Mey, 1980);

2) Mental activities that encompass thinking, imagining, remembering and problem solving (Allen, 1991);
3) Is part and parcel of decision making (Tang & Solomon, 1998).

The ensuing chapters involving methodology, the collection of data, analysis, results and conclusions are based on the assumption that relevance involves a cognitive process that is mediated by a system of categories based on user defined concepts for the purpose of resolving an information-related problem by making an evaluative decision.

<u>Measure, Measuring, Measurement</u>

1) The extent, dimensions or capacity of anything (Agnes, 1996)
2) The idea that relevance judgments form a stable and valid foundation on which to construct measures of retrieval performance is essentially an untested assumption (Harter, 1996);
3) The problem of measurement in information retrieval research is not that of measuring relevance, but of employing relevance as a measure or criterion of measurement (Ellis, 1996).

Many studies have intermingled a user's cognitive approach to judging with that of an IR system's ability to perform effectively. These are two separate and distinct actions, the latter is measurable and the former is not. This semantic distinction is a construct that girds the methodological approach to this investigation by creating a separation of the judgment process from the evaluation process. Judging, in this study, maintains its own cognitive content for the user.

Judge, Judging

1) To form an opinion about (Agnes, 1996)
2) The mental ability to distinguish relationships (Agnes, 1996)
3) There is a difference between the employment of a device to measure a physical process and the employment of human judgments (Ellis, 1996);
4) Relevance judging is a categorization process (Froehlich, 1991).

As further conjectured by Froehlich (1991) and incorporated into this investigation, relevance judging may be a step in the cognitive evaluative process leading to the resolution of an information need but it may not be the only step.

Evaluation

In the context of the above definitions, relevance emerges not as a measure and not as a judgment, but as an evaluative process consistent with a conceptual framework derived by Saracevic (1995), yet extends beyond his viewpoint that equates evaluation with performance. Evaluation as a cognitive process becomes an expression of importance, worth or value that may be converted into a measure, but in and of itself does not possess that quality. Difficulties have arisen when, during the execution of that evaluative process, users have not been provided with opportunities to be as discriminating as possible, but may have been required to be more discriminating than they are able (Tague & Schultz, 1989). A strict binary construction of value (relevant or not relevant) prevalent in IR system performance evaluations is the key example that has enjoyed prominence in the literature which may contrast profoundly with the actual cognitive conceptualizations of the user during his/her seeking, searching, retrieving and evaluating activities.

These judging, evaluating and measuring constructs and their differences are an important aspect of uncovering a more accurate assessment of the nature of user cognitive behavior in association with information retrieved and evaluated by a user in an attempt to resolve an information problem at hand. This approach could lay the groundwork for a greater understanding of how and why users may approach their information problems through a process of discriminating, differentiation and orientation that allows for the succeeding evaluation encompassing the worth and value of the items retrieved from an IR system.

Discrimination and Orientation

Cognition-related literature contains many discussions of how humans discriminate, differentiate and orient themselves to new information; however, many interdisciplinary investigations have suffered from a lack of integration into studies of human information seeking, searching, retrieving and evaluating behavior in information science. There have been several investigations both within and outside the discipline, however, that have contributed support to this study.

Wilson (1968) suggested early that conceptual representations of relevance should include three notions to avoid confusion. Those notions support the model of this investigation. Wilson's logical relevance could represent the user's first step in orienting to items retrieved from an IR system. His 'textual means to an end' could represent the aggregating synthesis that users employ to label or categorize their logical orientation to what has been viewed. This subjective aggregating synthesis by users has generally been the bane of most qualitative investigations because user's responses such as 'right on target' or 'not exactly' or 'interesting' are not always easily converted into performance measures. Wilson's third notion of relevance as satisfaction could represent a point in the process where the user decides what to do next.

Frants and Brush (1988) proposed a model of information need that looks at cognition as a process that involves categorization by identifying features and properties. MacMullin and Taylor (1984) indicate that the information continuum from a user perspective moves from specifying to connecting to orienting. The orienting process is what adds clarity and meaning to the problem in relation to the need. Janes (1994) recognized the user's use of overlapping concepts as orientations to a judgment process, but concluded that they must be just part of some unnamed larger entity. Park (1993) also understood that the evaluation process requires an orientation first to the factors underlying the relationship between the user's need and the retrieved item and then a second step that involves the choice. Park, however, was unable to operationalize a distinction between labels of orientation factors and labels for choice decisions.

Although Barry (1994) was able to conceptualize the cognitive discrimination or orientation process that users employ, she did not structure a methodology to disclose its nature. She did demonstrate, however, that decisions about relevance (evaluation) are not decisions about document representations (judgments). The importance of this distinction was lost, however, because the conceptual framework surrounding her treatment of relevance equated judgment to evaluation which led to the notion of a multi-dimensional approach based on user defined criteria for relevance.

This study postulates that user defined criteria are significant for formulating an orientation to an overarching level of judgment which can then lead to a decision of importance, value or worth. The defining criteria are only the starting point.

Bruce (1994) also took an approach to operationalizing user cognition as an orientation process by addressing the user's cognitive schema. Once again, however, the cognitive schema that user's employed were treated as variables engendering both judgment and evaluation through conversion to magnitude scaling. This approach provided a reflection of the cognitive schema employed by users, but not how the various schema orientations led to the magnitude estimates.

Saracevic (1996) also supported an orienting multi-stage approach to relevance when he defined relevance as a relationship by inference leading toward on intention within a context. Although Saracevic considered the process of inference and intention to be an overlapping manifestation of user cognition, this study considers them as distinct elements in the overall evaluative process that users implement to move toward information need resolution.

Ingwersen (1996) provides a tripartite model of information seeking by identifying the user's knowledge state, cognitive state and problem state in relation to resolving an information need. That conceptual framework also lends credence to this study which assumes that all three of those states serve to model, not just information seeking as a whole, but can also model human information coordinating behavior when items are retrieved, judged and evaluated by users.

Kreitler and Kreitler (1976) from the field of cognitive science posit that cognitive orientation occurs as a reaction to uncertainty about the meaning of a stimulus or about the meaning in a given situation. This construct integrates well with user judgments during states of uncertainty in a given situation involving an information need as they are exposed to items of information during the retrieval and evaluation stages of information gathering to resolve their specific information problem. Kreitler and Kreitler identify feature analysis as the key issue in this cognitive orientation process that includes three stages. The first stage is simply an alerting and sampling to increase the inflow of information from the environment as a whole with minimum prior filtering. This view was synthesized from prior studies by Pribram (1971) and Berlyne (1960). A further synthesis constructs a view that takes this alerting, sampling and minimally filtered information and presents it to a user as a stage I orientation intervention, i.e. an initial exposure to items retrieved resulting from users interacting with an IR system.

The second stage in this orientation process consists of multilevel hierarchical extraction and analysis of specific features ranging from simple to complex as suggested by Selfridge (1959) and Sutherland (1969) and supported by Kreitler and Kreitler. This second stage renders the information more specific or more general depending on the relationship of the features to each other. It is also recognized that the variety of possible labels attached to these features must be easily classifiable into meaningful dimensions in the cognitive realm which is consistent with Kreitler and Kreitler's viewpoint, i.e. are the retrieved items on topic, informative, useful, etc.

In this study, this second stage orientation process is examined in order to more thoroughly explore this concept of feature analysis in the context of user relevance judging and the meaningful classifiable dimensions (levels of relevance) that users are able to incorporate into that process. This approach also serves to separate the judgment process from the evaluation process.

The third and final stage of this orientation process is an active synthesizing of the feature analysis into an overall statement or elicitation of the orienting response. In this study, the orienting response is the region of relevance that the user cognitively synthesizes as an overall orientation statement about the relationship of the retrieved item features to the information problem at hand, i.e. relevant, not relevant, partially relevant, etc. This third stage represents the cognitive evaluation process that identifies the region of relevance of an item retrieved from an IR system.

When the cognitive orientation is completed with a resulting elicitation from the user a fourth stage can be developed to convert the elicitation into a measure for evaluating performance by the user and/or the IR system. Most research in information science to date has viewed stages two through four as a single step cognitive process where judgment, evaluation and measure equate to a single variable or conclude that the multidimensional aspects of this process are too elusive. Mizzaro (1998), for example, has posited this multistage process as components (topic, task and context) in the fourth dimension of his relevance model, but provides no description of how users take their topic, task and context and cognitively derive judgments and make evaluative decisions about those components.

Medin (1976) looked closely at an information processing view of discrimination learning specifically as a multi-stage process. In the first stage the stimuli must impinge on the senses. The second stage converts the sensory stimuli into a description of the functional stimuli as a process of cognitive selection into components, compounds or configurations depending on the theoretical framework. According to Medin, these stimulus descriptions are then entered into working memory where they are associated with working outcomes that act as input to the decision process. In the context of Medin's conceptualization, an item retrieved from an IR system could act as a sensory stimulus which is then cognitively converted into selective categories (judgment) that act as input to the decision process (evaluation).

Similarly, Hayes (1993) asserts that humans cognitively manipulate data to create information as a multi-level process consisting of selection, structuring and reduction. That viewpoint is consistent with the discrimination and orientation framework girding this investigation.

In the area of behavioral decision-making, Tversky and Kahneman's (1985) prospect theory, addresses the human orientation process as an evaluation of prospects. In prospect theory, outcomes are expressed as positive or negative deviations from a neutral reference point considered having a value of zero. When users retrieve items from an IR system, whatever their basis for judgment might be, the outcome for each item may be viewed and expressed as some individual or combination of positive or negative expressions. Under prospect theory, people generally evaluate outcomes in terms of minimal aspects because creating such a mode simplifies evaluation and reduces cognitive load. This may also apply when users judge and evaluate items retrieved from an IR system.

Degrees of Relevance

Kelly's (1955) psychology of personal constructs provides a historical foundation for this study in two aspects. First, it precedes a long line of support in the information science literature for the concept of 'degrees of relevance.' Second, it provides a framework of understanding for the complexity of relevance and its implementation as an evaluative tool in information science.

Kelly's construct of human personality imposes three corollaries that support this study. The first is his organization corollary posing that humans systematize their constructs by first arranging them and then by abstracting them further. Humans cognitively build ordinal relationships to form higher order constructs. Another important corollary documented by Kelly is his dichotomy corollary. Humans utilize dichotomous relationships in order to identify which aspect should be considered in situations involving choice. The value of this corollary is that it assumes a process that lends itself to binary mathematical methods of analysis. Thirdly, Kelly identifies a choice corollary that assumes whenever humans are confronted with making a choice that relative values are placed on the dichotomies under consideration thus structuring the decision made. That decision may be transitory or permanent depending on the situation at hand. The value of this discussion as the precursor for a review regarding degrees of relevance lies in Kelly's recognition that although a dichotomous relationship can be defined as black vs. white, i.e. white vs. not white or black vs. not black, the construct can be relativistic. Kelly is firm in stating that relativism is not the same as ambiguity. He asserts that dichotomous constructs can be built into scales that represent further abstractions of the separate dichotomous values. The assumption in this study is that dichotomous constructs do not act as the scale, but could lead to a higher order abstraction that could be built into scales.

In early discussions of the conceptual value of relevance as means for measuring retrieval effectiveness, Fairthorne (1958) expressed concern for the 'excluded middle' in reference to the use of only binary evaluation. Hillman (1964) in expressing the same concern was firm in his commitment that degrees of relevance must be considered in defining a weaker notion of relatedness in terms of documents, queries and index terms.

Goffman (1965) also recognized early that the prepositional function characterizing the relationship of a query to an item retrieved from a document set should be allowed to take on more than two values ("0" and "1") and was among the first, from an information science perspective, to insist that relevance should be permitted to take on any value. He argued that no finite valued logic is adequate for representing the IR process. This early insight, however, does draw attention to the fact that a multiplicity of "0" and "1" values may be an avenue for characterizing relevance relationships.

Katter (1968) was able to identify evidence that degrees of relevance are a function of the number of discriminably different values of the stimuli presented and by filling in such values would contribute to the range of values that exist between relevant and not relevant. This current investigation attempts to provide further clarity to this concept by introducing ways in which users cognitively identify discriminate values represented by information objects to invoke an evaluative response within a range represented by degrees of relevance, yet collectively aggregate into a region of relevance.

Cooper (1971) straddled the issue regarding relevance by identifying it as a concept still undefined, and recognized that a restricted definition from logical foundations of probability that admitted to no degrees of relevance may be too restrictive, but an issue that is difficult to resolve.

Although Foskett (1972) does not speak directly to the issue of degrees of relevance he confirms that pertinence on the part of the user is a pattern of thought in the reader's mind which is entirely supportive of this investigation which includes the development of a methodology to take a look at examples of possible user patterns of thought that equate with their evaluations of items retrieved from an IR system.

Wilson (1978) was explicit in setting out that users cannot sort out their beliefs into two neat categories of true and false. While arguing for an epistemology that includes information and misinformation, he argued for a full range of possibilities between those extremes. As did Maron (1978) when pondering that if relevance is a quantitative concept, then we should be allowed to speak in terms of degrees of relevance.

Bar-Hillel (1980) argued that people order information by its perceived degree of relevance. In the context of this investigation, the key word is 'perceived.' Degrees of relevance are subjective notions of understanding established in the minds of users as evaluative orientations to the information at hand.

Sperber & Wilson (1986) insist that intuitions of relevance are not about the simple presence or absence of relevance, but about degrees of relevance defined as a relationship between an assumption and a context. Most important to this study, however, is their suggestion that relevance is twofold in nature. First it is a classificatory concept and secondly, it is a comparative concept. That framework is interpreted to mean that user judgments may serve as an a priori classification process to the ensuing cognition surrounding comparative evaluation that derives some degree of relevance in the mind of the user. Sperber and Wilson lend support to this conceptualization by considering that classification and comparison are not one in the same, but reside together in the relevance realm of mental processing.

Eisenberg (1988) postulates that all definitions of relevance share the implication that relevance can be measured in terms of "how much" or degree, and therefore are able to be measured using magnitude estimation (interval scale measurements).

Gluck (1996) used a 5-point scale and Howard (1994) used a 13-point scale to identify degrees of relevance representative of evaluative importance of retrieved information to user needs. Combined with Janes (1994) 100 millimeter scale and Brooks' (1995) use of a 7-point scale, all direct attention to the fact that users have the capability of identifying varying degrees of incremental relevance. These incremental degrees of relevance act as an indication of how valuable or important an item is to their current information problem or need. A next step needed to bring these diverse studies into perspective is to compare the nature of the incremental differences in these degrees of relevance. This current study attempts to move in that direction.

Willis & Thom (1996) understood in their search for more effective approaches to recall that combining multiple judgments objectifies the variations between individual ideas of relevance. Their work continues to support the idea that degrees of relevance as an evaluation by the user are the result of preceding combinatorial judgments.

Brooks (1997) in his conceptual framework of the relevance aura states the notion that bibliographic objects exhibit varying degrees of relevance contingent on the user's perspective. The problem with Brook's approach, however, is that he considered topicality as the only basis for identifying the user's perspective.

Tang, Shaw and Vevea (1999) suggest that not only do users utilize varying degrees of relevance to make evaluative decisions about retrieved items, they express the greatest confidence when a 7-point scale is utilized for measurement purposes. They further express that there appears to be no improvement in user confidence levels when more that 7 categories of relevance are implemented in a measurement scale.

These previous studies tend to support that user cognition involves a process that extends beyond binary relevance (relevant/not relevant) when making evaluations of the importance of items retrieved from an IR system to help resolve an information problem at hand. Although the retrievalists also continue to look for (or deny) solutions to degrees of relevance in a variety of algorithmic approaches to probability, current trends are beginning to emerge with a focus on conditional logics to model the information retrieval process. Van Rijsbergen (1999) in a recent discourse on that topic, while still seeking the most appropriate ‘if-then’ hypothetical, introduces the concept of possible world semantics. His approach to this non-classical logic for IR remains a query to document operational relationship; yet this conceptual framework of possible world semantics may also yield applications in document to judgment to evaluation models of user cognitive processing during IR interactions. This study adopts and adapts Van Rijsbergen’s conjecture by positing that users create possible world semantics for judgment during interactions with IR systems. This possible world metaphor is introduced, discussed and operationalized in this study as levels of relevance.

Levels of Relevance

Belkin (1978) in leading to a discussion of ‘information as category’ states that different ideas of information exist at different levels and that an interdependence of levels generates a unifying general concept. Although that presentation by Belkin was not focused on user cognition during the evaluation of items retrieved from an IR system, the current investigation draws heavily on that notion as posited by Belkin as his description of the nature of information.

Park (1993) was able to discern that users’ relevance assessments involve multiple layers (levels) of interpretation in order to evaluate items retrieved for the resolution of an information problem.

Harter (1996) recognized that users bring ‘a great deal of personal baggage’ to the cognitive process of evaluating items retrieved from an IR system that extends beyond strict topicality considerations. His call for new evaluative measures represents a prime impetus for this study by looking at how some of these other considerations impact on user evaluations.

In a similar context, Barry and Schamber (1998) indicate two key assumptions in examining user cognition during evaluation. One assumption is that users base their evaluations on factors beyond the topicality of the information retrieved. The second assumption is that each individual does not possess a unique criteria set for making relevance judgments. Both of these assumptions, user cognitive assessment that extends beyond topicality and the belief in a core set of relevance concepts, is operationalized in this study.

Saracevic (1996, 1999) has introduced a taxonomy of relevance manifestations or levels of relevance that unifies an emerging understanding within the field of information science that categorizes how users discriminate or orient themselves toward items retrieved from an IR system during the process of seeking information. It is also recognized and acknowledged that the term ‘unifies’ may be a stronger interpretation than Saracevic intended with his ‘uneasy consensus’ position. However, the approach to this investigation is in agreement with his conceptual framework as a model of user cognitive manifestations that include topicality, pertinence, utility, motivation and system constraints as major categories that users consider during the judging stage of their overall evaluative processing of retrieved information.

Cosjin and Ingwersen (in press) in looking at the manifestations of relevance described by Saracevic, conclude that there is an interacting system of relevances on different levels and that different manifestations of relevance indicate different relations.

Those conclusions provide support for examining some of the different relations that exist as a result of these multiple manifestations of relevance and what those relationships may mean to the overall evaluative process that users cognitively invoke to resolve information problems and needs.

Complexity

This study, in part, looks at a possible approach for examining user cognition during the evaluative stage of information retrieval. To make any assessments, however, in a user's cognitive space requires a restriction to scope that can be treated empirically. One approach to measuring cognitive structure is to look at cognitive complexity.

Cognitive complexity was defined by Scott (1962) to be 'the number of independent dimensions-worth of concepts the individual brings to bear in describing a particular domain of phenomenon.' In this case, the particular domain of phenomenon is user relevance judging.

The notion of cognitive complexity has historical roots in the psychology literature to include Kelly's (1955) hierarchy of constructs, Zajonc's (1960) cognitive differentiation and Scott's (1969) dimensionality. Drawing upon a complexity framework, Linville (1982) posed that the complexity of our knowledge structures embodies an association of features that characterize an evaluative membership class. Utilizing terminology expressed by Abu-Mostafa (1986), information-based complexity has both an informational level and a combinatorial level. At the information level users might ask: Based on the information at hand, what is the intrinsic element of certainty (or uncertainty) that represents resolution (or irresolution) to the problem? At the combinatorial level the question becomes: What is the evaluative label that represents this combination of intrinsic elements as a solution to the problem at hand?

This investigation assumes that the features, elements, manifestations or levels of relevance that have emerged from the extensive body of relevance-related research in information science can be classified into a major, possibly exhaustive, set of categories that include systematic considerations, topicality, pertinence, utility and motivation. These categories have been defined by Saracevic (1999) as follows:

1) Systematic or algorithmic relevance: The relation between a query and the retrieved item of information in terms of the effectiveness of the system in returning what the user requested;

2) Topicality: The effectiveness of matching query terms to retrieved text;

3) Pertinence: The informative nature or meaningfulness of the retrieved text in relation to the information need;

4) Utility: The usefulness of the retrieved text in relation to the situation, task or problem at hand;

5) Motivation: The relation of the retrieved text to the intentions, goals or actions of the user related to the situation, task or problem at hand.

If users cognitively structure these intrinsic elements or levels of relevance into combinatorial levels of potential problem solution representations, how complex can these cognitive structures be? Based on the above-described 5-dimensional approach, users would be capable of 31 ($2^n - 1$) combinatorial cognitive judgment constructs ranging from a one-dimension construct up to a five-dimension construct. These possible combinatorial constructs may emerge and be implemented by the user as a cognitive judging process.

Searching for appropriate dimensions, attributes, features, elements, aspects, manifestations or levels of user cognition that model an understandable and useful framework of relevance has been conducted with many divergent methods. At this juncture, it is useful to introduce an apparent gap in past models of user cognition during relevance judging. That gap involves the conjunctive and disjunctive nature of human judgment behavior in choice situations. While the information science literature is somewhat lacking in this area, a considerable body of research surrounding human judgment, choice, problem solving and decision-making behavior lends substantial support to this study as cited and expanded upon in Chapter 3.

Conjunction and Disjunction

Svenson (1979) specified that, when placed in a decision-making situation, humans apply one or more cognitively established decision rules that may include choosing alternatives based on conjunctive and/or disjunctive aspects of the attributes associated with the decision needed.

As an operationalized definition, this study considers conjunctive aspects of attributes associated with a decision to be those that a user cognitively joins together as a result of their like values in helping to reach a needed decision as described by the levels of relevance identified above. Disjunctive aspects are those that, as a result of their disparate values, are also implemented cognitively to reach a decision.

If these definitions are applied to the levels of relevance that users employ to make relevance judgments, it becomes evident that a 5-dimensional cognitive space as posited above, actually exists as a 10-dimensional positive/negative cognitive judgment space. As an extension of the five levels of relevance listed above, the negative nature of those levels must also be considered as aspects of user judgment, i.e. not what I wanted from the system, not on topic, not pertinent, not useful, not a cause for further action. These five negative levels of relevance now contribute an additional 31 combinatorial possibilities to the user's judgment space considering, once again, that anywhere from one to five of these disjunctive levels can be invoked to make the judgment.

This model of human judging as a cognitive process is fairly parsimonious, yet enables a meaningful framework that contributes to the cognitive complexity surrounding user relevance judgments and how these judgments could impact evaluations represented by regions of relevance. More fully explained, users make judgments about items retrieved from IR systems as well as other resources. Those judgments include positive conjunctive associations at various levels (i.e. on topic and pertinent) and similar negative conjunctive associations (i.e. not on topic and not useful). However, users also appear to make disjunctive judgments represented by both positive and negative associations (i.e. not on topic, but informative and useful; or, on topic but not useful). The cognitive complexity associated with such judgments also requires investigation because it contributes an additional 180 [$3^n - 2(2^n) + 1$] combinatorial possibilities within the user's judgment space that have mostly been ignored or suffered from difficulties in operationalizing appropriate methodologies. That number of combinatorial possibilities is based on the five levels of relevance in a 10-category conjunctive/disjunctive space as described above. Assuming Saracevic's categories are a reasonable representation, users are capable of discriminating or orienting to five levels or manifestations of relevance when making judgments of retrieved items in an attempt to move toward resolution of an information problem or need. That conceptualization derives $3^n - 1$ combinatorial possibilities, which in a 5 level cognitive space equals 242 possible judgments. Evidence that users cognitively implement this range of possible judgments would contribute to a better understanding of the incremental degrees of relevance and the regions of relevance that users utilize during their evaluations of retrieved items.

In the language of models and processes that describe human cognitive behavior in judgment situations, this study takes a holistic approach in contrast to a dimension-wise approach (Wallsten, 1980). A holistic approach to classifying user judgments during IR interactions enables the examination of possible conjunctive and disjunctive thresholds that may contribute to user cognitive frameworks for making evaluative decisions of relevance. This approach contrasts with most dimension-wise approaches that have sought to isolate the aspects of user judgment that contribute the most to user decisions of relevance. While both of these approaches to investigation have merit, holistic strategies are preferred to dimension-wise methods if the dimensions of the alternatives under investigation are interdependent (Rosen & Rosenkoetter, 1976). A primary assumption in this investigation is that system considerations, topicality, pertinence, utility and motivation are interdependent at the point when users judge items retrieved from a search.

Synthesis and Issues for Resolution

An underlying issue for this investigation is whether user judgments are the same as user choices in evaluating the regions of relevance of items retrieved from an IR system or other resource to resolve an information-related problem. Einhorn and Hogarth (1981) make a clear distinction between these concepts by defining judgment as a deliberative reasoning process for evaluating evidence, while choice is the resulting evaluation. While somewhat murky as semantic descriptions, the key difference is that judgments are aids to choice. That distinction is important to this investigation because it could contribute to a synthesized model of the user evaluative process that extends beyond simply selecting a measured representation of relevance on some sort of scale.

Another issue, integral with the above, is the considerable research effort put forth by Fidel and Soergel (1983), Schamber (1991, 1994), Wang and Soergel (1993), Barry (1994), Wang (1994, 1997) and Barry and Schamber (1998) in defining criteria for relevance. In modeling a user evaluative process of information retrieval, the nature of these relevance criteria will require further discussion. Bateman (1998) indicated that these factors or criteria could influence user judgment, which, from the perspective of this investigation, requires an explanation of where these factors may fit into an evaluative process model of user interaction with an IR system.

A further issue that emerges from this investigative approach is the IR system development issue. Can binary algorithmic approaches to IR system design cope with the variety of user judgments that yield a range of inferences surrounding evaluations of relevance? Through this study, the dichotomous nature of user judgment behavior, not the levels of judgment themselves, may help in developing a "1's" and "0's" binary approach to user relevance evaluation processes.

Preliminary Studies

Partial Relevance

Prior investigations by this author regarding levels and regions of relevance provide a supporting foundation for the methodology and design incorporated into this study. While extensive research efforts have been put forth to describe the nature and manifestations of relevance in relation to user interactions with IR systems, few speak in terms of effects and predictability. One difficulty that has slowed research progress in that direction has been the inability to adequately describe and explain user cognition between the extremes of high relevance and non-relevant evaluations for items retrieved from an IR system.

Although user studies have not ignored this relatively unexplored middle region of relevance as a cognitive framework for user judgment and evaluation, most have ultimately collapsed their data by considering partially relevant evaluations as equivalent to relevant evaluations in measuring retrieval effectiveness (Saracevic, 1988; Su, 1994; Gluck, 1996). Spink and Greisdorf (1997), however, were able to identify an aspect of this 'fuzzy' region called partial relevance that distinguishes it from the extremes. Users, particularly users with little subject knowledge searching an information problem for the first time, tended to use items considered partially relevant as a 'springboard' for changing or redefining their information problems. Spink, Greisdorf and Bateman (1998) reported that finding as a positive correlation between partially relevant items retrieved and changes in users' information problems during the early stages of their information seeking process.

User Knowledge

User shifts or movements to a different cognitive plateau resulting from partial relevance evaluations points to an implied logic where users make relevance evaluations at the extremes (relevant or not relevant) based on what they know; whereas, partial relevance evaluations stem from what users don't know. Robins (1998) in his synthesis of prior work by Ingwersen (1996) lends support to this position in his discussion of user shifts of attention in relation to what they don't know (problem state) and what they do know (cognitive state). If users are required to make judgments about information items retrieved from an IR system, the resultant is derived from the user's cognitive state (what they know). The succeeding evaluation, once having made such judgment(s), is an estimation of the importance or value of the retrieved item, i.e. its degree of relevance, derived in conjunction with the user's problem space (what they don't know).

Support for this conceptualization has been provided by Spink, Greisdorf and Bateman (1999) in an investigation of user relevance evaluations resulting from mediated searching sessions. Findings from that investigation furnished evidence that successive searching on an information problem tends to decrease the number (based on the percentage of items retrieved and evaluated) of partially relevant evaluations while increasing the number of relevance evaluations at the extremes. If this 'learning by searching' conceptual framework is valid, then user cognition related to the informativeness of an item retrieved from an IR system may be more critical than the topicality of that item.

Recognizing a conceptual differentiation of user judgment from user evaluation, Greisdorf and Spink (1999) continued to develop an operationalized framework for describing and defining the middle range of relevance to include partially relevant and partially not relevant evaluations by users. Findings from that investigation indicated that middle values within a range of relevance evaluations, regardless of scale, are generally classified by users based on what is missing, as well as what is present, in the item or document. That indication of multiple features, both positive and negative, contributes to an understanding of relevance consisting of more than dichotomous choices at the extremes. A further investigation into the nature of these positive and negative levels of relevance is the focus of this current study.

Overlap

An additional concept posited as a result of the preliminary work reported above is represented in the current study. When users aggregate item or document features to make judgments about them, the same aggregated judgments may represent different regions of relevance to different users. Spink and Greisdorf (1999) found that overlapping judgments across regions of relevance provide an indication that items retrieved and judged in the middle regions (partially relevant and partially not relevant) could have an inferred meaning to users that extends beyond strict adherence to query match (topicality) alone. The range of these inferences may vary across regions of relevance, thus having an impact on traditional measures of precision and recall. An examination of these inferences is part of this current work in response to the original research questions posed.

Summary

To date, most investigations of user information seeking, searching and retrieval behavior in relation to relevance evaluations have treated the relevance judging process as a search for the positive orientations that users make to retrieved information in attempts to resolve information problems and needs. This study attempts to provide evidence that user negative orientations are just as valuable in providing a clearer explanation of user cognition during stages of relevance evaluation.

Literature in the interdisciplinary fields of information science, psychology, communications and cognitive science provide significant support to this line of investigative study. Looking at relevance judgment and evaluation as a problem solving decision-making process that encompasses these other lines of research and theories may further clarify relevance and its use as a tool for measurement in the evaluation of user performance, IR system performance and the effectiveness associated with interactions of the two.

CHAPTER 3

METHODOLOGY AND RESEARCH DESIGN

Introduction

Users categorize or classify cognitive levels or aspects of the information sought prior to searching for that information. Upon retrieving items for investigative purposes in resolving their information problems, users invoke judgments based on those categorized aspects or levels of relevance that may then be cognitively aggregated into some single valued region of relevance in relation to the problem and/or in relation to other items already examined.

A model of user evaluation would lend clarity to the process stages that users go through to reach an evaluative decision of relevance. To model that user evaluation requires a combination of underlying assumptions. To answer the research questions presented in Chapter 1, this study proceeds by looking at both a synthesized model of user cognition as an evaluative process and the identification of means for investigative study at specific stages of the model.

Underlying Assumptions

Three underlying assumptions contribute to the development of the methodology used in this investigation. The first assumption is that human evaluation is a cognitive process that involves judgment and decision-making that engenders a starting point and an ending point with cognitive activity intervening as stages or steps in the process. Prior investigations lending support to that assumption include the work of Newell and Simon (1972), Montgomery and Svenson (1976), Medin (1976), Kreitler and Kreitler (1976), Montgomery and Willen (1999), and Svenson (1999).

The second assumption is that during interactions with an IR system, humans determine the relevance of the information retrieved as an evaluative process that includes judgment and decision-making as explicated by assumption one. Relevance judging as an evaluative process has been supported and discussed in the literature by Saracevic (1995).

The third assumption is that relevance judging as an evaluative process can be equated with cognitive decision-making and can be examined through judgment analysis and decision research methodologies.

A Model of the User Evaluation Process

Simon (1955, 1972) recognized that there is a point in the cognitive process of evaluation where the user stops and does something next. However, a question important to this study is where the evaluation process begins. Contrary to most contemporary thinking in information science, the concept of an anomalous state of knowledge (Belkin, 1978) may not only be the initiation point of user information seeking activities, it may also represent a starting point in user evaluations. That is, user queries only consist of what users know about what they do not know. During evaluation, however, the initiating ASK, after a user reviews an item retrieved from an IR system, may consist of both what the user knew they did not know and what the user did not know that they did not know.

This could be considered a more robust ASK. A more robust or enhanced ASK may begin anew the process of evaluation for the user.

The next step, derived from the supporting literature, is some cognitive categorization, discrimination or orientation of beliefs that incorporates the criteria from which users choose to make judgments about the items being viewed after retrieval in relation to the current ASK. The difficulty is that there is no clear evidence that differentiates how these criteria are used for the purpose of querying the system from how these criteria are used during evaluation. The premise underlying this investigation is that criteria-based IR system queries may or may not be the same criteria for evaluation of the items retrieved. For example, criteria such as authority or currency may be used to initiate IR system algorithms for retrieval based on user's beliefs that such information will meet their needs. Once an item is retrieved, the user may or may not continue to look at those criteria as a basis for moving to another stage in the evaluation process.

That next stage or step in evaluation assumes, based on the previously identified studies of Kreitler and Kreitler (1976) and Medin (1976), that these various criterion-based user beliefs may be aggregated into a higher level of cognition, which then are weighed against the evidence presented to the user by the retrieved item under scrutiny. These higher levels of relevance identify the needs of the user in relation to evaluating the item. This concept is important because these needs may not be directly related to the information problem of the user, they may only relate to the immediate evaluation of the item retrieved in relation to the user's overall base of knowledge. This distinction is significant because it may disengage the current perceived needs of the user from the system delivering results on the query-generating initial ASK. These aggregated levels of relevance may never be stated by the user, but are recognized as the implicit level of judgment that leads to a value or region of relevance in the mind of the user. The methodology devised for this study looks at this stage of user cognition during evaluation because its transparent nature seldom emerges in studies of user relevance.

Once these 'needs' are identified, the user has achieved a cognitive stopping point in the evaluation process. However, as an imposed condition of examination, developed mostly for measurement purposes, users have been asked to express aggregations of need using some form of relevance scale in order to develop measurement criteria for retrieval effectiveness (i.e. precision, recall, fallout). This cognitive exercise inserted into the evaluation process is a stage that may not be implicit in the user's cognitive evaluation process, but is an imposed condition that the user must deal with on terms imposed by the experimental design. This next step may not occur in real life evaluative situations of user interactions with IR systems, it may only be an evaluative artifact imposed on the user for experimental purposes. In one sense, this study is no exception, however, it explicitly recognizes this artifact in attempting to model a user's evaluation process.

When these stages are completed in relation to an item being viewed by the user, either the initiating ASK has been moved toward resolution, the originating ASK remains unresolved or a new ASK is evoked. In each case, the evaluation begins anew with another retrieved item under examination by the user.

This evaluation model could be represented as a five-stage process consisting of: 1) starting with some information need, 2) a base of cognitively generated criteria-laden features of the retrieved item for judgment purposes, 3) a meaningful aggregation of those features based on analysis of the retrieved item into some overarching level(s) of relevance or importance to the user at the time the item is examined, 4) an assigned value to those levels of relevance as some measure of the item's importance, i.e. the degree or region of relevance to the user in relation to the information problem at hand, and 5) stopping the evaluation process or viewing the next item retrieved with a new, modified or unchanged information need.

This model, as depicted in Figure 3.1, represents user cognition during relevance judging and evaluation during interaction with an IR system. The task in this study is to look at these stages in a manner that illuminates understanding of the user in relevance judging situations.

START → **FEATURES** → **AGGREGATED NEED** → **ASSIGNED VALUE** → **STOP**

Stage 1 | Stage 2 | Stage 3 | Stage 4 | Stage 5

Figure 3.1: Model of User Cognition as an Evaluative Process

A refined version of the model can be constructed in the context of conceptual frameworks developed in information science as represented in Figure 3.2.

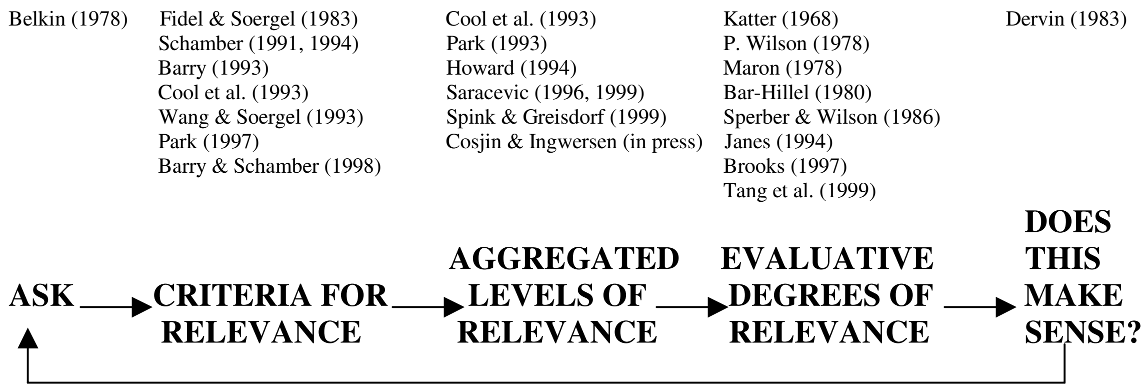

Figure 3.2: A Model of User Relevance Judging and Evaluation during IR interactions

This model enables investigating the original research question that asks whether users maintain distinct consistent evaluative heuristics for determining regions of relevance. Research has already provided an indication that users' criteria for relevance are inconsistent and not reducible to single level variables. In addition, a criterion, factor or feature approach would involve an unmanageable manipulation of variables based on Fidel and Soergel's (1983) 200 factors, Turner's (1992) 160 information attributes, or Schamber's (1994) consolidated list of 80 relevance factors. If users aggregate their needs as an approach to reducing cognitive load, as suggested in Figure 3.1, then looking at possible aggregations of user needs in the form of levels of relevance could be an approach for answering the research question. Determining whether users, in fact, aggregate their needs into measurable levels of relevance requires a design framework capable of detecting thresholds for evaluation.

Design Framework

Representative Design

Research on human decision-making has generally included conceptual frameworks that could be considered parallel but distinct from information science approaches for describing and explaining human relevance judging as a cognitive activity. Approaches that span these disciplines in the area of relevance judging by IR system users could draw attention to alternative methodologies that could generate useful data and associated analyses for a greater understanding of user relevance perspectives in information retrieval environments.

Humans in judgment situations generally perform under constraints from many sources, both real and perceived, encompassing problems with accurately measuring information, task constraints that include time and information availability, and factors of emotion and social influences. These issues and concerns have been addressed by Saracevic (1995) in information science and corroborated in the judgment and decision-making literature by Cooksey (1996) to include 'representative design' as a useful approach for analyzing human cognition in judgment situations.

Historically, representative design has involved the study of human behavior under naturally occurring multi-variable mediating conditions in the task environment as an observed, not controlled, investigation. Pioneered by the works of Brunswick (1952, 1955, 1956) and expanded upon by Hammond (1955, 1975), the Lens Model conceptual approach to representative design fits the requirements of this investigation.

Lens Model Representation Design

The Lens Model approach to investigative study as originally envisioned by Brunswick (1952) and adapted for the study of human judgment by Hammond et al. (1975) is shown in Figure 3.3.

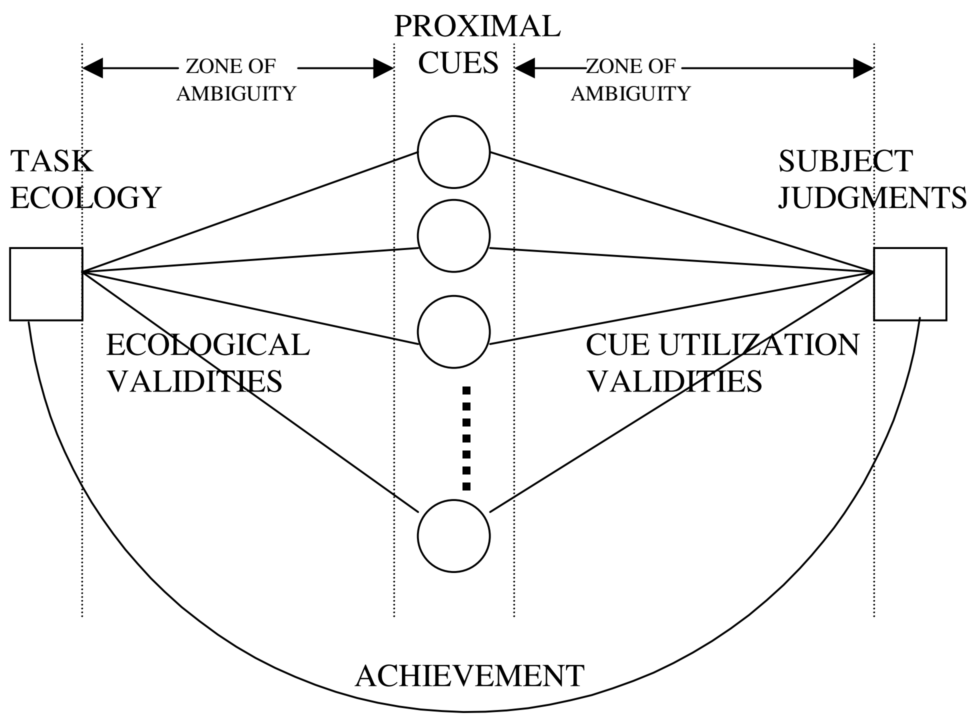

Figure 3.3 Lens Model for the study of human judgment (adapted from Cooksey, 1996)

In the context of relevance judging, the operationalized terminology indicated in Figure 3.3 can be defined as follows (as adapted from Cooksey, 1996):

Ecological Criterion is a representation of a user's cognitive or ecological task at a point in time that the judgment is made. Ecological criteria in user relevance judging situations could include, but not be limited to, a users' state as an inference or relation of:

1) Item retrieved to initial query
2) Item retrieved to initial problem
3) Item retrieved to modified query
4) Item retrieved to modified problem
5) Item retrieved to new query
6) Item retrieved to new problem
7) Item retrieved to unrelated problem

Unstated, or uncontrolled, ecological task criteria could lead to confounding results in approaching studies of user relevance in information retrieval situations which has led to the methodology envisioned for this investigation.

Proximal or Surface Cues are any textual, numeric, verbal, graphical, pictorial or other sensory stimuli available to the judge (user) in forming a judgment of the item retrieved from an IR system. For example, relevance criteria such as authority, currency, comprehensiveness and/or style could act as proximal cues to user judgment.

Ecological Validities represent any system for quantitatively summarizing cue emphasis by the user to model an ecological criterion, i.e. topicality is a proximal cue that can model the ecological criterion of query to item retrieved in terms of relevance; however, topicality may or may not model the ecological criterion of an item retrieved to the initial problem in terms of relevance.

Cue Utilization Validities represent any system for quantitatively summarizing cue emphasis for predicting judgment process outcomes. In relevance judging investigations, the predictive nature of proximal cues (criteria for relevance) has been elusive because of their numbers. For example, examining the predictive power of Schamber's (1994) summary of 80 relevance factors as an operationalized investigative study would be, at the least, cumbersome and time consuming, and at the most, impossible to coordinate effectively.

Subject Judgments are a user's appraisal of a cue with respect to some dimension of interest. These judgments are usually quantitative and may be scaled in terms of a specific metric or as a categorical judgment. This aspect of the model makes it suitable for relevance investigations where user judgments can take on interval, ordinal or nominal scale characteristics.

Zone of Ambiguity is the cognitively entangled relationships that a user must cope with in order to make a decision. The region is ambiguous because relationships of cause and effect take place in both directions, i.e. between task and cues, as well as between cues and judgment. Hammond et al. (1975) referred to this ambiguous nature of the judgment process as the 'causal texture' of these varying relationships. The value of the zone of ambiguity concept to this study is the recognition of the existence of these entangled relationships accompanied by the ability to move the focus of the investigation beyond them.

Achievement in the Lens Model represents the focusing aspect of the design by defining the relationships among ecological task, cue generation, and judgment of the cues in relation to the task. In Lens Model representations, however, the researcher may or may not know the user's task ecology. That gives strength to the model, especially within the realm of the current study. Often in studies utilizing relevance judgments, users' task ecologies remain hidden or confounded by the operationalized nature of the investigation.

The application of a Lens Model approach to this study is derived from its expandability in a variety of operational settings. Two specific extensions of the model are utilized to examine user relevance thresholds as an approach to the research questions posed. The first extension of the model is called the n-system design (Cooksey, 1996) approach that takes into account many judges (users) with the purpose of capturing and comparing a number of relevance judgment representations. The value of this n-system approach is that it allows for differing cue validities (relevance criteria) as well as differing consistencies among users, i.e. differing user dimensions of interest. A model of n-system design is shown in Figure 3.4.

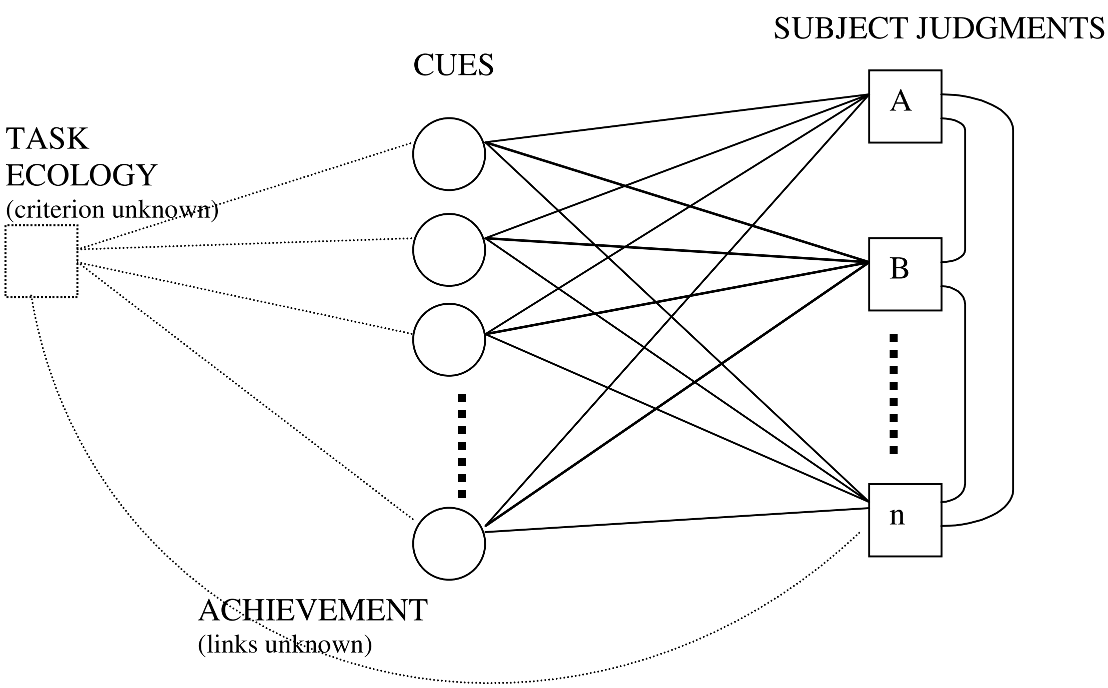

Figure 3.4 n-System Lens Model design for comparing several judgment systems (adapted from Cooksey, 1996)

This n-system design approach allows the exploration and examination of cues, features, aspects and dimensions of relevance that may be impacted by user task, knowledge and situation without knowing their specific nature. It further allows an examination of how multiple cues can influence multiple judgments within and between multiple users. These 'within and between' user judgments can then be studied to determine their relationships to the cues under investigation. This extension of the investigative model for the study of relevance judging stems from conclusions derived by Hammond et al. (1975) as described in Cooksey (1996) that state:

1) People do not generally describe their judgment reasoning accurately;
2) People do not apply the same reasoning consistently;
3) Generally, people only use a small number of cues to make a judgment;
4) It is difficult to understand another person's cognitive reasoning surrounding judgment just through observation and verbal explanations;
5) The identification of additive organizing principles may be adequate to describe judgment processes.

The second extension of the Lens Model is the hierarchical judgment approach. It extends the n-system model to include first order cues that can generate first order judgments that can act as second order cues leading to second order judgments. This extension of the Lens Model, as shown in Figure 3.5, will serve as the foundation for approaching relevance thresholds.

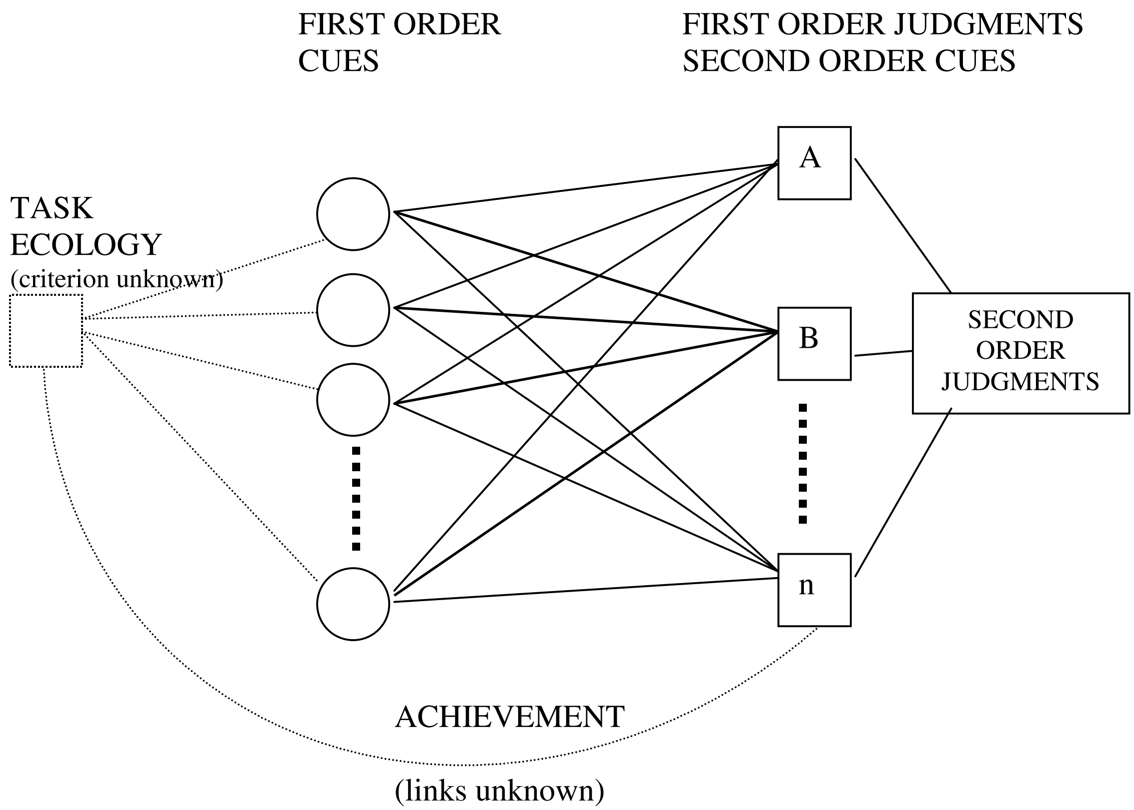

Figure 3.5 Hierarchical n-system model where judgments at one level serve as cues for the next level with the task ecology criteria unknown (adapted from Cooksey, 1996)

Hierarchical judgment modeling allows a set of aggregated cues that act as first order judgments to be examined in terms of their relationship to second order judgments that relate to the existing task ecology. This investigative approach to relevance allows users to model dimensions of interest surrounding items retrieved from an IR system where the researcher is without knowledge of or unable to control the underlying task or the initiating (first order) cues. For the purpose of this study, elements of the n-system hierarchical model have been labeled with terms appropriate for relevance judgment investigation as shown in Figure 3.6.

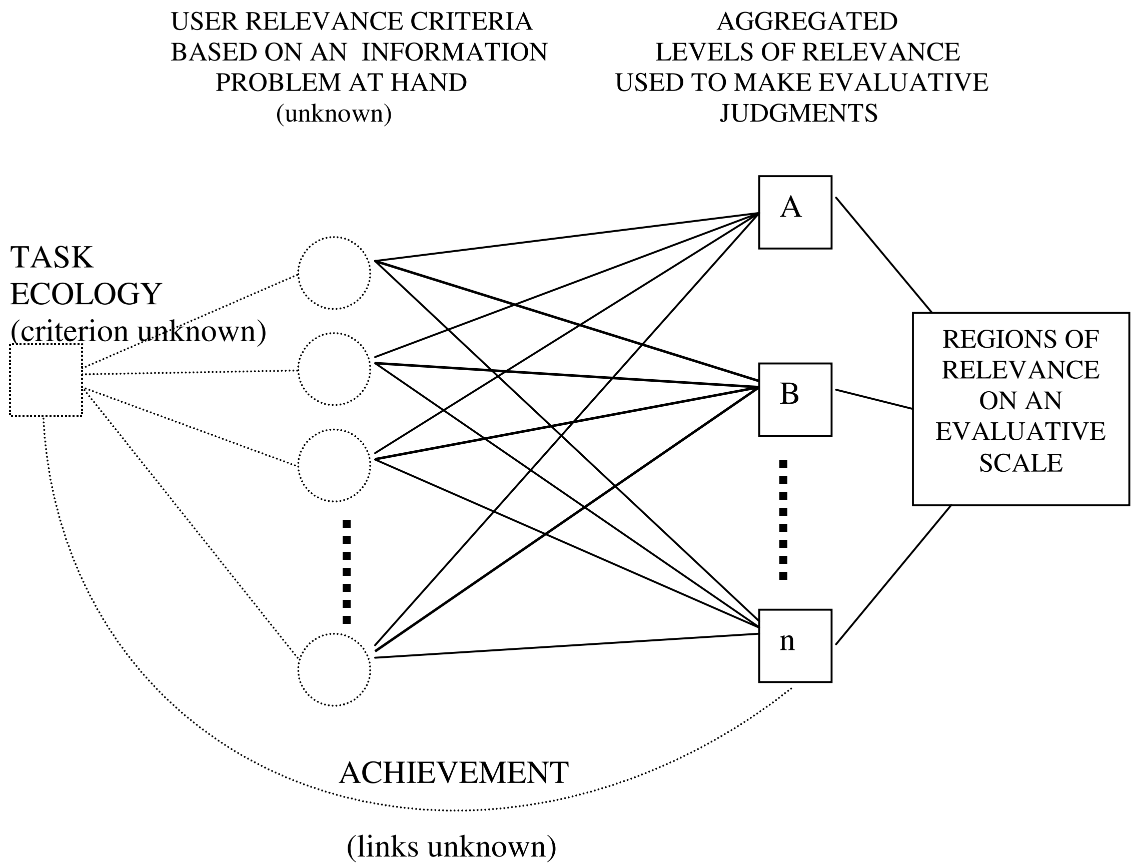

Figure 3.6_ Application of a hierarchical n-system lens model to study end-user relevance judging where the task and specific initiating cue criteria are unknown.

Hypotheses

The hypotheses that frame the following analyses are introduced here now that the primary research questions, defining concepts, prior research and a model supporting the investigation have been presented. The approach is exploratory in the sense that the emergent hypotheses stem from a cursory view of the data and prior investigations by the author. The hypotheses can be stated as:

1. End-users of IR systems use a multi-stage cognitive process that exhibits characteristics of a Lens Model representative design for evaluating retrieved

items. Those stages of cognitive processing will exhibit the following characteristics:

(a) Consideration, which will be demonstrated if end-users choose evaluative levels of relevance (systematic, topicality, pertinence, utility, motivational) in relation to an information problem at hand in categories that significantly exceed or fall short of the probabilities of chance selection;

(b) Differentiation, which will be demonstrated if end-users choose positive and negative aspects of these evaluative levels of relevance in a manner that significantly exceeds or falls short of the probabilities of such selection by random chance; and,

(c) Aggregation, which will be demonstrated if end-users combine evaluative levels of relevance to describe regions of relevance (not relevant, partially not relevant, partially relevant, and relevant) in a manner that significantly exceeds or falls short of the probabilities of such selections as random chance.

2. Positive and negative aspects of item characteristics (levels of relevance) will demonstrate clustering into different regions of relevance in a manner that significantly exceeds or falls short of the probabilities of random selection;
3. The effects of systematic, topicality, pertinence, utility and motivational levels of relevance in terms of their positive and negative aspects and the end-user clustering of those aspects will be demonstrated by the manner in which these

levels significantly contribute to end-user evaluative responses that exceeds or falls short of the probabilities of random selection.

These hypotheses are related to the Lens Model design for studying human judgment and the research questions in the following manner:

1(a) Consideration: Underlying cues or criteria for human judgment are cognitively processed to determine which cues or criteria will be used to structure how an object is judged. That same process takes place when an end-user must take those judgments and move to a higher order judgment stage represented by some type of evaluative response.

1(b) Differentiation: After the cognitively appropriate cues are determined for judging, each cue is weighted in order to render a judgment about the object. That same process takes place when an end-user must determine the positive or negative nature of their judgments in order to render an evaluative response.

1(c) Aggregation: After consideration and the weighting of judgment values in relation to an object under investigation, the consolidation or aggregation of the positive and/or negative judgments structure an evaluative response to the overall value of the object. For an end-user this overall value can be expressed as some region of relevance that is either self described or imposed by experimentation.

2. Positive and negative clustering: As end-users structure the value or importance of an item retrieved from an IR system, the clustering of positive and negative aspects of their underlying judgments create cognitive conjunction or disjunction for making evaluative responses. If all of the aggregated judgments or either positive or negative,

conjunction is in evidence. If the aggregated judgments are a ‘mixed bag’ of positive and negative judgments, disjunction is in evidence.

3. Levels of relevance (judgments) in relation to regions of relevance (evaluation): The characteristics for judgment (levels of relevance) under consideration in this study (systematic, topicality, pertinence, utility, motivational) may or may not play a particular cognitive role for end-users for determining the value or importance of items retrieved from an IR system in relation to an information problem at hand. These relationships, if they exist, are under investigation as an integral part of this study.

For this exploratory research, and in consideration of the nominal characteristics of the data, a Chi-square two-step approach will be used for analysis. First, each hypothesis will be tested for significance by using the Chi-square statistic for modeling a view of end-user cognition during the evaluation process associated with item retrieval from an IR system to resolve an information problem at hand. Any rejection of the null will be considered as significant beyond mere chance observation. Second, each significant Chi-square statistic will be further analyzed to model end-user evaluative processing by looking at the standardized residuals (Hinkle, Wiersma & Jurs, 1994) associated with the significant Chi-square. This approach is assumed appropriate for this exploratory research.

A Priori vs. A Posteriori Judgment Decomposition

Developing a methodology for analyzing user relevance judgments by looking at the cues that stimulate those judgments can be structured as a decomposition of the cues prior to making the relevance judgments (a priori) or after users have made their relevance judgments (a posteriori). These approaches to judgment analysis are discussed in Arkes and Hammond (1986). For purposes of data collection and analysis, this study uses an a posteriori approach by asking users to make relevance judgments on a categorical scale of relevance and then decompose those judgments into the cues utilized for making those judgments. The decomposed cues are the levels of relevance represented as first order judgments and second order cues in the model shown in Figure 3.6.

Users, Objects and Tasks

The selection of users as relevance judges encompasses issues of face validity and construct validity in relation to the task being operationalized. With regard to this study, users were asked to retrieve items from an IR system in relation to a specific information-related problem they themselves had an interest in resolving. The purpose of that approach is to provide a high level of face validity to the task that users are being asked to perform. By asking each and every user to retrieve items regarding an externally posited search question could reduce face validity and confound results because users might not take the task at hand seriously.

The construct validity of this investigation emerges if users, regardless of their level of knowledge, can decompose cues (levels of relevance) into the dimensions of importance that contribute to their relevance judgments. If users are able to cognitively map the cues available to the relevance judgments made, high construct validity can be assumed.

Items retrieved for evaluative purposes by users will be considered as information objects regardless of their form and format. This investigation is not an examination of surrogate versus full text effectiveness. It will identify whether users are able to use judgment cues (levels of relevance) to reach a decision about regions of relevance regardless of the form and format of the information object.

The focus of attention in this investigation is user ability to identify (decompose) information objects into their dimensions of importance. This cognitive exercise seeks to clarify the positive and negative levels of relevance that users utilize to make relevance judgments. These relevance thresholds are represented by conjunctive and disjunctive aggregations of relevance levels. Conjunctive aggregations express all positive or all negative relationships between the problem and the item retrieved. Disjunctive aggregations express mixed positive and negative relationships between the problem and the item retrieved. These conjunctive/disjunctive thresholds are examined with regard to their descriptive, explanatory and predictive nature.

Data Collection

A self-reporting structure served as the means for users to document their relevance judgments and the relevance cues representing those judgments. Self-reports may be considered as valid reconstructions of decision-making when the influential factors are plausible, are included in a priori theories, and there are no plausible non-influential factors (Nisbett & Wilson, 1977). The levels of relevance utilized as user cues for decomposing their relevance judgments are considered plausible based on the prior research identified in the literature review reported in Chapter 2, and the representative design afforded by a Lens Model approach to judgment analysis provides an underlying theoretical framework that would allow self-reports as a functional and valid approach to this study. Previous literature (Bradburn, Rips & Shevell, 1987; Sudman & Bradburn, 1982) and the Lens Model approach to judgment analysis provide a robust theoretical framework for asserting that self-reports are functional and valid in the context of this study.

Limitations

By considering user relevance judgments as a cognitive evaluation process both the depth and breadth of this investigation will be encumbered by the following limitations:

1. *External Environment:* Initial criterion or cue selection that may serve to anchor user judgments to the query formulated for IR system input is not a focus of this study. This condition is considered a variable that may mediate the secondary cues (levels of relevance) under investigation and are thus accounted for by their impact on secondary cue selection by the user.
2. *Central Processing*: Individual combinations of user knowledge and experience that contribute to judgment and evaluation are ignored in this study except to the extent that all users have been identified as having the capabilities necessary to conduct an IR system search. Those capabilities include problem identification, query formulation, system query parameters, item retrieval and item review for evaluative purposes in seeking to resolve an information problem at hand.

3. *Feedback*: No feedback mechanisms are included in this study except to the extent that any item judged by a user can have a feedback effect by providing new information that may impact succeeding judgments and evaluations. Considering that this is a naturally occurring phenomenon in IR system retrieval situations, no effort has been implemented to control this mediating variable during data collection.

Although these limitations exist in information seeking, searching and retrieval environments, the literature and Lens Model suggest a user's ability to identify, select and aggregate appropriate cues for making relevance judgments, including measurable values for such judgments, should not be experimentally hindered by these limitations.

Summary

In both real life situations and experimentally imposed environments, users bring a great deal of 'personal baggage' to the evaluation process associated with information seeking, searching and retrieval (Harter, 1996). Attempts to delineate, distinguish and segregate single level variables that compose an individual's 'personal baggage' in any given situation involving information needs remain inconsistent.

The approach to this investigation is to recognize that a user's cognitive posture in a given situation does not necessarily require a decomposition into its least common denominators. How those least common denominators are aggregated into a categorized higher level of cognition, however, could provide avenues for identifying and explaining individual differences in IR situations more succinctly.

CHAPTER 4

DATA ANALYSIS

Characteristics of the Data

The conceptual approach to the data analysis assumes that human cognition functions as an aggregation of features that define meaningful concepts. From the perspective of users of IR systems that are seeking information to resolve an information need, an analysis of contributing features and the resulting defining concepts are an integral part of furthering an understanding of relevance as a multi-faceted cognitive evaluative activity.

Experimentally imposed choices are critical to an understanding of how users employ cognitive choice when making judgments about the value or importance of information they retrieve from an IR system to resolve an information problem. As discussed in Chapter 2, humans choose to consider or not to consider the mental representations that will accomplish their objectives, they utilize both positive and negative aspects of what they consider, and they aggregate their considerations to achieve a manageable representation of both their judgmental and evaluative needs. Those aggregations may be conjunctive in the sense that they are all positive or all negative, or they may be disjunctive in the sense that the meaningful aggregation contains both positive and negative features. For this investigation users have been allowed to consider or not consider the variables presented, the variables have been presented in both their positive and negative forms, and the resulting conjunctions and disjunctions have been identified for analysis.

Implementing the Lens Model

Any analysis of human judgment should include appropriate tasks for obtaining judgments that accurately reflect the decision process of the judge (Cooksey, 1996). The Lens Model provides a representative design and set of parameters that reflect user interactions with IR systems. Those parameters, defined as task familiarity and task congruence, act as an underlying typology for user judgment in this investigation.

A dimension of task familiarity carries the implication that all of the judges (users) participating in this investigation have made these types of judgments before in real life and have been targeted for this research based on prior experience in resolving information needs by utilizing information seeking, searching, retrieving, judging and evaluating cognitive processes. For example, the graduate students participating in this study have conducted searches on prior information problems, retrieved items for examination and performed judgments to evaluate their importance or value to the problem at hand.

A dimension of task congruence (Adelman, 1981; Miller, 1971) implies that the labels attached to the conceptual cues (levels of relevance) presented to users in the investigation correspond logically within the task required (judging and evaluating items retrieved from an IR system) and the criterion measure (regions of relevance) used to understand the relationships between the cues and the criterion measure. The first order judgments (levels of relevance) presented to users were structured around Saracevic's five consensus manifestations of relevance (systematic, topicality, pertinence, utility and motivation) to keep the number of conceptual cues presented to users within Miller's (1956) range of five to nine chunks of information in working memory. Most judgment analysis research employs between 2 and 20 cues (Stewart, 1988). An attempt has been made with this study to insure that the judgment cues are representative of those concepts that are potentially salient to users when making judgments and evaluations surrounding an information need.

To maintain congruity among users in the interpretation of the task environment a worksheet (Appendix A) and an instruction set (Appendix B), given to the users prior to their search and retrieval tasks, presented a uniform procedure for recording the conceptual judgment cues (levels of relevance) and the evaluative measure (regions of relevance). This provided a means for maintaining congruity among users in their interpretation of the task environment.

Worksheet for Data Collection

Users retrieved items from an IR system for review, judgment and evaluation based on their own information problems, no demands for any specific number of items to be retrieved and with no interference by the researcher. Then each user completed a worksheet that identified the region of relevance that the item represented in relation to the specific information problem. Then each item was identified by the most important reasons for making that evaluation. The structure of the worksheet provided check boxes for users to record their evaluations and reflect a posteriori on their reasoning for such evaluations, as shown in Figure 4.1.

ITEM	JUDGMENTS [Check one box only]				LEVELS OF RELEVANCE [check box(s) most important to your relevance decision]									
	NR	PNR	PR	R	S	T	P	U	M	NS	NT	NP	NU	NM
(each worksheet accommodated up to 20 items for judgment and evaluation)														

Figure 4.1 Data Collection Worksheet

An instruction set accompanied each worksheet giving clarifying directions and definitions as follows:

(1) For each item retrieved in this search, identify its relevance to your information need by placing a check in the box (check just one) that identifies whether the item is considered not relevant [NR], partially not relevant [PNR], partially relevant [PR] or relevant [R];

(2) Identify the level of relevance by checking the box(s) that indicate the most important reasons for making the relevance decision as you did. These levels are defined as follows:

[S] Systematic: The item retrieved was in a form/format that meets my information need;

[T] Topicality: The item retrieved was on the topic/subject I requested;

[P] Pertinence: I believe the item retrieved is/will be informative;

[U] Utility: The item retrieved is/will be useful in resolving my current/or a future information need;

[M] Motivational: The item retrieved will/may cause me to take other action(s) now that I have this information;

[NS] Systematic: The item retrieved was NOT in a form/format that meets my information need;

[NT] Topicality: The item retrieved was NOT on the topic/subject I requested;

[NP] Pertinence: I believe the item retrieved is NOT/will NOT be informative;

[NU] Utility: The item retrieved is NOT/will NOT be useful in resolving my current/or a future information need;

[NM] Motivational: The item retrieved will NOT/may NOT cause me to take other action(s) now that I have this information.

In the context of this investigation of user judgment and evaluation and the implementation of a Lens Model approach, it is necessary to emphasize that users were not given definitions for the categorical values imposed as regions of relevance. Thus, the ensuing analysis is able to determine if there is any relationship between levels of relevance and regions of relevance based strictly on how users cognitively define those values as individuals. That notion is important for determining if there are distinctive relationships that define a middle range of relevance between the extremes of relevant and not relevant evaluations by users. The analysis has been designed to capture these relationships (if they exist). Understanding these relationships would help clarify a model of user cognition during the relevance evaluation process.

Data Corpus

Analysis was conducted on data included in relevance worksheets (Figure 4.1) from 32 users judging and evaluating 1432 *items* retrieved from IR systems encompassing a variety of databases. *Item* is used to describe the range of retrieved entities that included citations, abstracts and full text documents. The users can be profiled as a convenience sample of graduate students enrolled in the School of Library and Information Science at the University of North Texas. The data was collected over time (1997 and 1998) to aggregate a pool of responses consistent with the parameters for appropriate judgment analysis that suggests a minimum 5 to 1 ratio (Cook, 1976) of cue profiles (sets of judgments by a single user) to every cue utilized. With systematic, topicality, pertinence, utility and motivation considered as the cues under investigation, then the 5 to 1 ratio would suggest a minimum of 25 users for the investigation.

Data Analysis for Modeling User Judgments

The a posteriori judgments made by users have been treated as the independent variables in this study because the operationalized Lens Model implies that these judgments acted as the cause for evaluating the items retrieved in the categories identified by the worksheets. The user evaluations defined by the categorical regions of relevance (NR, PNR, PR and R) are treated as dependent variables as a result of their positioning as second order judgments in the Lens Model.

The judgment gap or zone of ambiguity in the Lens Model that exists between the first and second order judgments is mediated by user cognition. For this investigation, user cognition is defined as a 3^n process. The '3' represents the cognitive options of cue consideration in the positive sense (S, T, P, U, and M), cue consideration in the negative sense (NS, NT, NP, NU, and NM), and no cue consideration (where the fill in boxes for a particular level both positive and negative were left blank). The 'n' represents the number of variables under consideration for making a second order judgment. For this study, n=5 to represent the systematic, topical, pertinent, useful and/or motivational cues deemed important to the user in making a decision about the relevance of an item retrieved.

The collected data can be represented by a series of partitioned contingency tables suitable for non-parametric statistical analysis since the variables, both dependent and independent, are presented at a nominal level of measurement.

Hypothesis 1(a): Consideration vs. Non-consideration

In order to determine if users consider whether or not various parameters contribute to the evaluation process, there must be a set of parameters. These parameters could be established a priori or a posteriori by the researcher or the user. For this investigation a set of five judgmental parameters, defined as levels of relevance (systematic, topicality, pertinence, utility, motivational), was imposed by the researcher a priori. The first step in the analysis of the data was to determine if those parameters fit the model of relevance posed by the investigation. That is, do users, at least some of the time, use a multi-step aggregating process to determine relevance. Table 4.1 presents the combined decisions of the 32 users in evaluating the 1432 items retrieved based on whether there was consideration or non-consideration of the imposed parameter for making their judgments.

	Systematic	Topicality	Pertinence	Utility	Motivation	Totals
Considered	469	919	884	825	430	3527
Not Considered	963	513	548	607	1002	3633
Total	1432	1432	1432	1432	1432	7160

Table 4.1 Judgment parameter consideration by users

A closer statistical analysis of the contributions made by each parameter or level of relevance illuminates a more substantive model of user judgment processes. A cursory observation of the data indicates that users on the whole were just as likely to consider these parameters (49%) as not (51%). While the total population of users in this study utilized these parameters on almost a 50-50 basis, a significant chi-square statistic (X^2 = 624.60; with X^2_{cv}= 9.49 at α =.05 with df=4) produces a set of standardized residuals (Hinkle, Wiersma & Jurs, 1994) that can be used to identify the significant contributions to this modeling of relevance levels. Examining only *consideration* versus *non-consideration*, only topicality (std. residual = 8.04), pertinence (6.72) and utility (4.50) appear to provide significant contributions, whereas systematic (-8.90) and motivational (-10.37) considerations were significant by contributing less to the model than could be expected by chance.

The early implication is that topicality, pertinence and utility appear to be the cognitive judgments that users are most likely to implement when making evaluations of retrieved items in an attempt to resolve an information problem. However, further analysis is required to determine how these parameters of judgment are positioned by users when making evaluations of relevance. Likewise, further analysis is required to determine if systematic and motivational levels contribute to modeling relevance based on differentiations by region of relevance.

Hypothesis 1(b): Differentiating Levels of Relevance

Table 4.2 presents an expanded tabulation that gives the frequencies of the considered items based on user judgments in a positive or negative sense as they were evaluated by region of relevance.

		Not Relevant NR	Partially Not Relevant PNR	Partially Relevant PR	Relevant R	Total
Systematic	Positive	16	50	93	202	361
	Negative	79	19	10	0	108
	NC	446	150	187	180	963
Topicality	Positive	79	58	154	313	604
	Negative	257	43	15	0	315
	NC	205	118	121	69	513
Pertinence	Positive	6	25	137	313	481
	Negative	270	105	28	0	403
	NC	265	89	125	69	548
Utility	Positive	1	14	96	264	375
	Negative	320	75	55	0	450
	NC	220	130	139	118	607
Motivation	Positive	2	16	38	138	194
	Negative	180	42	13	1	236
	NC	359	161	239	243	1002
Total		2705	1095	1450	1910	7160

Table 4.2 Tabulation of Worksheet Data Collected (NC = not considered)

The 7160 total responses were derived from user judgments based on the five judgment cues (systematic, topicality, pertinence, utility, and motivation) including three aspects of those cues (considered positive, considered negative or not considered by the user) multiplied by the 1432 items judged (5 x 1432 = 7160). Analysis of the data in table 4.2 provides a significant chi-square statistic (X^2 = 2901.77; X^2_{cv}= 57.53 at α =.05 with df=42) which prompts further investigation of the standardized residuals (Table 4.3) to determine which levels of relevance contribute toward modeling user evaluations (regions of relevance). Either a positive value of the judgment cue, a negative value or a non-consideration of the cue could be viewed as being important to a user in evaluating a retrieved item.

Items Judged Relevant [R]		Items Judged Partially Relevant [PR]		Items Judged Partially Not Relevant [PNR]		Items Judged Not Relevant [NR]	
First Order Judgments	Standardized Residuals	First Order Judgments	Standardized Residuals	First Order Judgments	Standardized Residuals	First Order Judgments	Standardized Residuals
U(pos)	16.39	P(pos)	4.01	P(neg)	5.52	T(neg)	12.65
P(pos)	16.30	T(pos)	2.86	T(nc)	4.46	U(neg)	11.50
M(pos)	11.99	M(nc)	2.53	U(nc)	3.86	M(neg)	9.62
T(pos)	11.97	S(pos)	2.33	M(neg)	0.98	P(neg)	9.54
S(pos)	10.77	U(pos)	2.30	U(neg)	0.74	S(neg)	5.98
M(nc)	-1.49	T(nc)	1.68	M(nc)	0.63	S(nc)	4.31
U(nc)	-3.45	U(nc)	1.45	S(neg)	0.61	P(nc)	4.03
S(nc)	-4.80	P(nc)	1.33	P(nc)	0.57	T(nc)	0.80
S(neg)	-5.37	M(pos)	-0.21	S(nc)	0.22	U(nc)	-0.62
T(nc)	-5.80	S(nc)	-0.57	S(pos)	-0.70	M(nc)	-1.00
P(nc)	-6.38	S(neg)	-2.54	T(neg)	-0.75	M(pos)	-8.33
M(neg)	-7.81	U(neg)	-3.78	M(pos)	-2.51	T(pos)	-9.88
T(neg)	-9.17	M(neg)	-5.03	T(pos)	-3.58	S(pos)	-10.31
P(neg)	-10.37	P(neg)	-5.93	P(pos)	-5.66	U(pos)	-11.82
U(neg)	-10.96	T(neg)	-6.11	U(pos)	-5.72	P(pos)	-13.04

Table 4.3 Standardized residuals allocated to each region of relevance where S=systematic; T=topicality; P=pertinence; U=utility; M=motivational; and where (pos)=positive consideration; (neg)=negative consideration; (nc)=not considered.

Standardized residuals are considered significant at values in excess of +2, while negative residuals at values in excess of -2 are considered as contributing less to the model than could be expected by chance. Values in between +2 and –2 are considered as non-contributing factors due to their probability being no greater than chance selection by the user. A graphical representation of the residuals presented in Table 4.3 in descending order provides an indication of how these levels of judgment contribute to a model of each region of relevance. Each graph depicts an individual region of relevance (relevant, partially relevant, partially not relevant, not relevant).

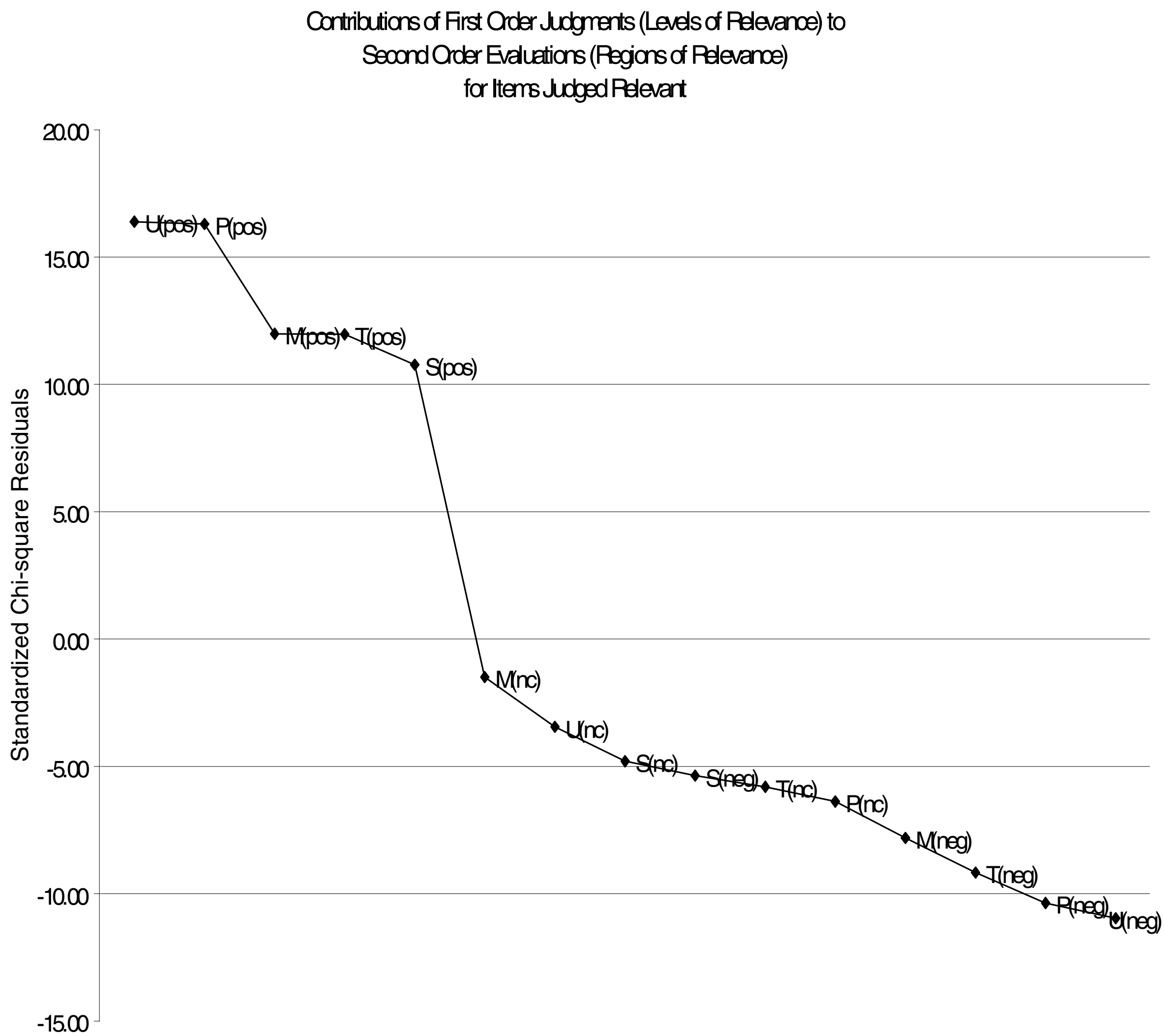

Figure 4.2 Judgments contributing to a model of relevant evaluations

The first graph in Figure 4.2 identifies the standardized residuals associated with all items evaluated by users as relevant. Since the values of all positive levels of relevance under investigation exceed +2 they would all appear to contribute to a model of relevance at the extreme normally considered as the region of highly relevant items. This would imply that all of the levels of relevance utilized in the investigation appear to model what users consider to be relevant in the positive sense when viewing, judging and evaluating items retrieved from an IR system, called the region of relevant items.

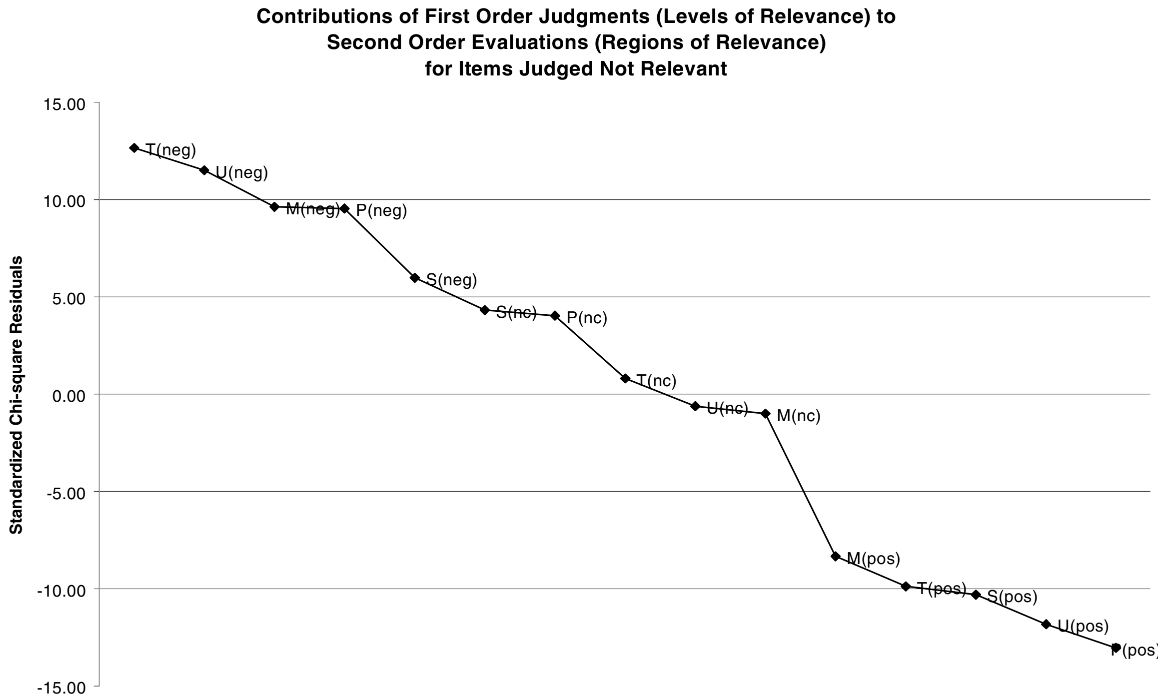

Figure 4.3 Judgments contributing to a model of not relevant evaluations

In the region of not relevant evaluations the residuals identify significant contributions from all five levels of relevance in the negative sense. Contributions to a model of non-relevance also come from non-considerations of systematic and pertinence judgments which could be interpreted to mean that these levels may not be as important to users as a non-relevant judgment cue. For example, if a full-text document is retrieved, the user most probably will not consider its form or format as a reason for discarding it [S_{nc}]. If it is not full-text, the negative aspect of systematic considerations [S_{neg}] could be used as a judgment contributing toward de-selection by evaluating it as not relevant. The converse would be true of pertinence. If a full-text document is retrieved, the user might consider it not pertinent [P_{neg}] for one reason or another; but, if it is not full-text it is many times difficult for the user to even make a judgment about pertinence. In that case, pertinence is not likely to be considered [P_{nc}] as an evaluative feature of non-relevance.

At this point, the intuitively obvious appears to be empirically confirmed at the extremes of relevance. That is, from a dichotomous perspective, the extreme regions of relevance appear to conform to prior findings in relation to the positive nature of relevant evaluations and the negative nature of not relevant evaluations.

Exposing Middle Regions of Relevance

The middle regions of relevance as categorically defined in this study as partially relevant and partially not relevant yielded significant residuals as contributions to a model of relevance as shown in Figures 4.4 and 4.5.

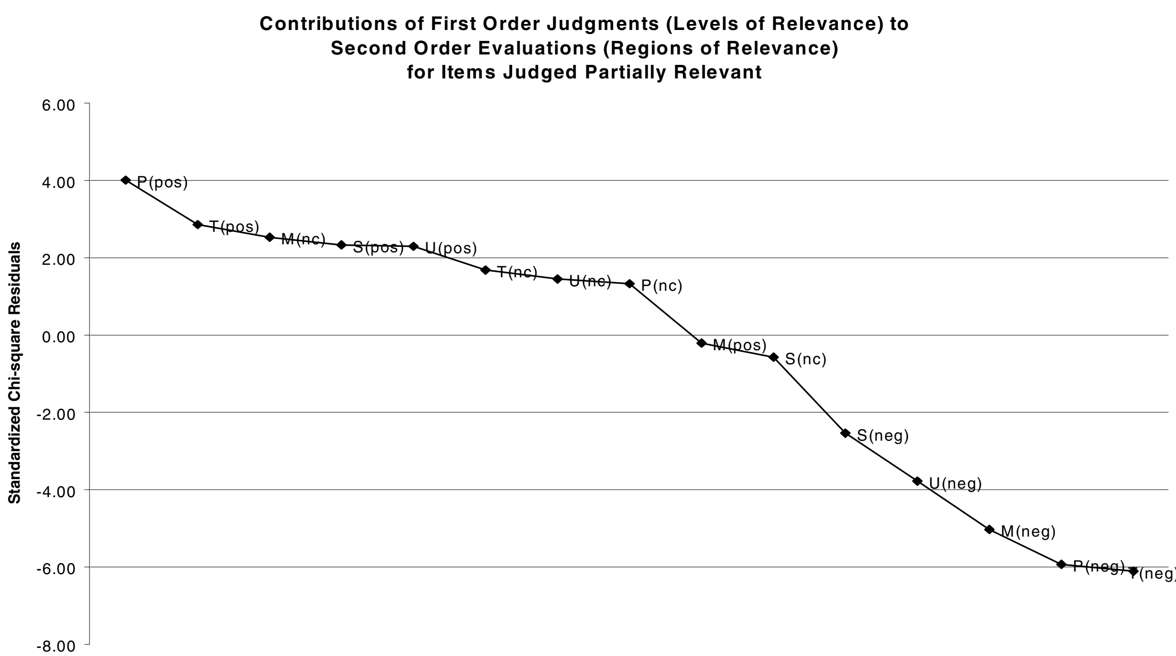

Figure 4.4 Judgments contributing to a model of partially relevant evaluations

In the region of partially relevant evaluations, the residuals identify significant contributions from only four levels of relevance in the positive sense [pertinence (+4.01), topicality (+2.86), systematic (+2.33), and utility (+2.30) considerations]. The non-consideration of motivational judgments (+2.53) also appears to contribute to a model of partially relevant evaluations which could mean that some form of causative action is not a judgment cue (under the search circumstances of this study) that users generally associate with items evaluated as partially relevant.

In the region of partially not relevant evaluations (Fig. 4.5), the residuals identify significant contributions from only one level of relevance represented by pertinence in the negative sense (+5.52). The non-consideration of topical (+4.46) and utility (+3.86) judgments also appears to contribute to a model of partially not relevant evaluations. If partially not relevant items represent some inference about a retrieved item that is not sufficient to reject it (evaluate it as not relevant), then topicality and utility may not be the judgment cues for making that type of evaluation, while pertinence may be the cue used for that evaluation. The implication is that the item may be on topic, but that aspect of the item is not the most important cause for a user evaluation of partially not relevant.

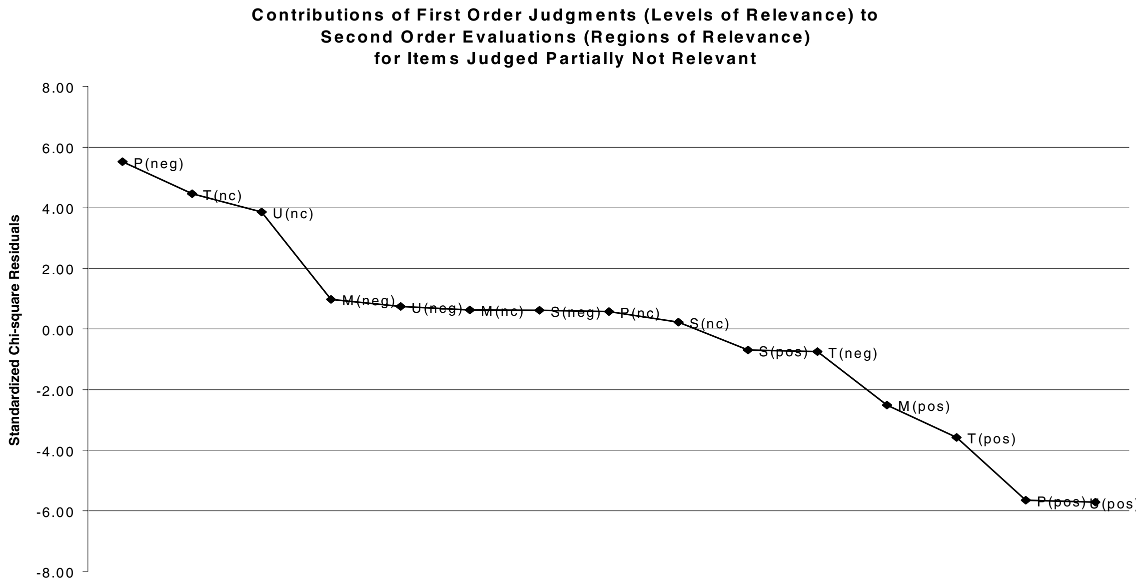

Figure 4.5 Judgments contributing to a model of partially not relevant evaluations

Hypothesis 1(c): Analysis of Cognitive Aggregation

The above analysis offers some insight into the levels of relevance (first order judgments) that users consider important in evaluating the relevance (region of relevance) of an item retrieved from an IR system. Additional analyses of the cognitive aggregation of those levels into a meaningful evaluative response were required as a next stage in this investigation.

If users make judgments based on an aggregation of features, the choice behavior in this study is represented by 242 possibilities ($3^5 - 1$) as shown in Appendix C. These possibilities are represented by 5 levels of relevance aggregated in a conjunctive positive sense ($2^5 - 1 = 31$), the same 5 levels of relevance aggregated in a conjunctive negative sense ($2^5 - 1 = 31$), plus the same 5 levels considered disjunctively including both positive and negative aspects [$3^5 - 2(2)^5 + 1 = 180$] These 242 possible combinations of first order judgments help the user structure the second order judgments of relevance as an evaluative response. Those possible combinations, based on the five levels of relevance utilized in this study, include the user's ability to aggregate their judgments as 1, 2, 3, 4 or 5 levels of relevance to describe an evaluative region of relevance (not relevant, partially not relevant, partially relevant, or relevant) as shown in Appendix D. Table 4.4 provides a summary of user judgment aggregations based on the region of relevance to which that aggregation was cognitively assigned for the 1432 items retrieved.

Relevance evaluation based on:	NR	PNR	PR	R	Total
1 level of judgment	235	106	103	72	516
2 levels of judgment	130	57	68	57	312
3 levels of judgment	52	15	89	75	231
4 levels of judgment	61	10	13	69	153
5 levels of judgment	63	31	17	109	220
Total	541	219	290	382	1432

Table 4.4 Aggregated judgments by region of relevance

A chi-square analysis of the data provided in Table 4.4 indicates a significant result (X^2 = 230.51; X^2_{cv}= 21.03 at α =.05 with df=12) prompting a further analysis of the standardized residuals to determine which aggregations contribute significantly to a model of relevance regions. A summary of those residuals is provided in Table 4.5.

Relevance evaluation based on:	NR	PNR	PR	R
1 level of judgment	2.87	3.05	-0.15	-5.60
2 levels of judgment	1.12	1.34	0.61	-2.88
3 levels of judgment	-3.78	-3.42	6.17	1.70
4 levels of judgment	0.42	-2.77	-3.23	4.41
5 levels of judgment	-2.21	-0.46	-4.13	6.57

Table 4.5 Chi-square standardized residuals

Significant values (>+2) provide an indication that single cue judgments are most likely to occur in the not relevant (std. residual =2.87) and partially not relevant (std. residual =3.05) regions of relevance as users evaluate items retrieved from an IR system. Three levels of judgment are more likely to occur when users evaluate items as partially relevant (std. residual =6.17). In the relevant region, it appears that users incorporate an aggregation of 4 to 5 levels of relevance (std. residuals =4.41 and 6.57). in order to determine if an item is actually relevant in terms of their current information need. These results point to a view of user heuristics during retrieval evaluation that are consistent with Blair's (1990) concept of futility versus utility where the user constantly strives to retrieve useful documents while struggling with the de-selection process as efficiently as possible. That is, as positive aspects are sought and considered, once a negative feature is exposed through user cognition, it is most likely that the retrieved set of items is reduced (not relevant evaluations) or segmented (partial vs. highly relevant) for additional consideration depending on the task or problem at hand.

Hypothesis 2: Conjunctive/Disjunctive Modeling of Relevance

Further partitioning of the collected data reveals the cognitive reflection of users in relating positive and negative judgment cues to an evaluative relevance process. If users consider relevance as a judgmental process to determine the importance or value of an item retrieved from an IR system, then some cognitive filtering(s) of cues probably act as the mediating judgments that lead to an evaluative decision about the items retrieved. This study has chosen the positive/negative cognitive dichotomy as the filtering variable of investigation to determine if user judgments tend to follow some logical sequence in discerning meaningful values of relevance within a range of relevance choices.

Table 4.6 presents a partitioning of the 1432 items judged and evaluated by the users' conceptual positioning of conjunctive versus disjunctive judgments in relation to the regions of relevance to which they were allocated. In Table 4.6, positive conjunction is defined as some aggregation of only positive judgment cues (S, T, P, U, M) to describe a particular region of relevance. Negative conjunction is defined as some aggregation of only negative judgment cues (NS, NT, NP, NU, NM) to describe a particular region of relevance. An aggregation of judgment cues can consist of any single or combination of cues from one to five as implemented by users to make a relevance evaluation. Disjunctive judgments are defined as any combination of judgments that include an aggregation of both positive and negative judgment cues. Disjunctive aggregations consist of any combination of cues from two to five elements.

Relevance evaluations based on:	NR	PNR	PR	R	Total
Positive Conjunction	1	20	191	381	593
Negative Conjunction	449	114	5	0	568
Disjunctive Judgments	91	85	94	1	271
Total	541	219	290	382	1432

Table 4.6 Conjunctive and Disjunctive Relevance Judgments

A significant chi-square statistic calculated from Table 4.6 (X^2=1299.38; X^2_{cv}= 12.59 at α =.05 with df = 6) generates standardized residuals that identify the significant contributions that model relevance based on conjunctive and disjunctive judgment cues. Positive conjunctive cues appear to contribute significantly in the regions of partially relevant (std. residual = 6.47) and relevant (std. residual = 17.72) evaluations, whereas negative conjunction appears to contribute to the partially not relevant (std. residual = 2.91) and not relevant (std. residual = 16.00) regions. Disjunctive judgment cues appear to contribute to modeling relevance only in the middle regions of partially relevant (std. residual = 5.28) and partially not relevant (std. residual = 6.77) evaluations.

Partitioning relevance into three major regions of relevant, partially relevant (including partially not relevant evaluations) and not relevant evaluations can model parsimony missing from the above analysis. An all-inclusive middle range of relevance evaluations can be modeled with a similar analysis of the standardized residuals resulting from a significant chi-square statistic (X^2 = 1120.02; X^2_{cv}= 9.49 at α =.05 with df=4) derived from the partitioned contingency table shown in Table 4.7.

Relevance evaluations based on:	NR	PNR / PR	R	Total
Positive Conjunction	1	211	381	593
Negative Conjunction	449	119	0	568
Disjunctive Judgments	91	179	1	271
Total	541	509	382	1432

Table 4.7 Conjunctive/Disjunctive Judgments in the Middle Regions of Relevance

This analysis points to a model of relevance that implies users make positive conjunctive judgments (std. residual = 16.00) to evaluate items relevant and negative conjunctive judgments (std. residual = 17.72) to evaluate items not relevant. Disjunctive judgments appear to model a middle range of partial relevance (std. residual = 8.42) that fills the gap between relevant and not relevant evaluations. These significant residuals can be identified in Table 4.8.

Relevance evaluations based on:	NR	PNR / PR	R
Positive Conjunction	-14.90	0.02	17.72
Negative Conjunction	16.00	-5.83	-12.31
Disjunctive Judgments	-1.12	8.42	-8.38

Table 4.8 Conjunctive/Disjunctive Standardized Residuals

Hypothesis 3: Relating Levels of Relevance to Regions of Relevance

From the Lens Model perspective, the above analyses point to a model of user cognition during relevance judging and evaluation that asks these questions:

1. If users have a tendency to utilize single level (Table 4.5) aspects of systematic, topical, pertinent, useful and motivational considerations to de-select and discard

items retrieved by evaluating them as not relevant, which level is most likely to be considered by the user for making that decision?

2. If users have a tendency to utilize single level (Table 4.5) aspects of systematic, topical, pertinent, useful and motivational considerations to evaluate items as partially not relevant, which level is most likely to be considered by the user for making that decision?

3. If users have a tendency to utilize three levels of relevance (Table 4.5) that include aspects of systematic, topical, pertinent, useful and motivational considerations to evaluate items as partially relevant, which levels are most likely to be considered by the user for making that decision?

4. If users have a tendency to utilize four or five levels of relevance (Table 4.5) that include aspects of systematic, topical, pertinent, useful and motivational considerations to evaluate and select items as relevant to their information needs, which levels are most likely to be considered by the user for making that decision?

The above questions can be addressed, since each item judged and evaluated reflected an aggregated combination of levels as requested by the instructions to each user. A table of these aggregations based on the significance of the residuals in Table 4.5 is provided in the following tables.

Not Relevant Evaluations based on one level of judgment	
Judgments	Items
NT	109
NP	59
NU	54
NS	11
NM	1
	234*

Table 4.9 Single level judgments evaluated by users as not relevant

*Actual not relevant judgments by users totaled 235, a single [T] judgment has been eliminated from the analysis by designating it as an anomaly which could bias the results of the analysis of this table only.

Of the 541 items evaluated by users as not relevant 235* (43%) of those evaluations incorporated single levels of relevance as the most important reason for making the relevance decision as indicated. A summary of the standardized residuals resulting from a significant chi-square (X^2 = 159.16; X^2_{cv}= 9.49 at α =.05 with df=4) associated with Table 4.9 indicates only one significant residual value (9.09) for items considered to be not on topic [NT].

Partially Not Relevant Evaluations based on one level of judgment	
Judgments	Items
NT	20
NP	60
NU	13
NS	6
NM	4
M	2
T	1
	106

Table 4.10 Single level partially not relevant judgments

Of the 219 items evaluated as partially not relevant, 106 (48%) were evaluated using only a single level of relevance as indicated in Table 4.10. A summary of the standardized residuals resulting from a significant chi-square ($X^2 = 173.08$; $X^2_{cv} = 12.59$ at $\alpha = .05$ with df=6) associated with Table 4.10 indicates only one significant residual value (11.53) for items considered to be not pertinent [NP].

Partially Relevant Evaluations based on three levels of judgment	
Judgments	Items
T/P/NU	32
T/U/NP	19
S/T/P	13
S/T/U	5
S/P/U	4
S/P/M	4
S/T/M	3
T/P/U	3
S/U/M	2
S/M/NT	2
S/P/NU	1
P/M/NT	1
	89

Table 4.11 Three level judgments evaluated by users as partially relevant

Of the 290 items evaluated as partially relevant, 89 (31%) judgments used three levels of relevance (Table 4.11). Standardized residuals resulting from a significant chi-square ($X^2 = 131.99$; $X^2_{cv} = 19.68$ at $\alpha = .05$ with df=11) indicate three significant residual values that include items considered to be [T/P/NU] on topic, pertinent, but not useful (9.03), items considered to be [T/U/NP] on topic, useful, but not pertinent (4.25) and items considered to be [S/T/P] in the right form/format, on topic, and pertinent (2.05).

Relevant Evaluations based on four or five levels of judgment	
Judgments	Items
S/T/P/U/M	108
S/T/P/U/NM	1
S/T/P/U	59
S/T/P/M	5
T/P/U/M	5
	178

Table 4.12 Four and five level judgments evaluated by users as relevant

A summary of the standardized residuals resulting from a significant chi-square (X^2 = 248.85; X^2_{cv}= 9.49 at α =.05 with df=4) associated with Table 4.12 indicates two significant residual values that include items considered to be [S/T/P/U/M] in the right form/format, on topic, pertinent, useful and motivational (12.13), and items considered to be [S/T/P/U] in the right form/format, on topic, pertinent and useful (3.92).

Summary of Results

Within this study users of IR systems cognitively judged and evaluated the results of their search and retrieval efforts. The analyses support the concepts of multiple judgment cues, aggregation, and conjunctive/disjunctive evaluation as cognitive aspects of user heuristics when items are retrieved from an IR system for the purpose of resolving an information problem at hand. The conjunctive and disjunctive nature of evaluative judgment cues is demonstrated when users are required, either by a self-imposed process or by an experimentally induced procedure, to identify the importance or value of the retrieved items. There is support for the appropriateness of the researcher-imposed parameters of this study, though no exclusivity of judgment cues was sought or indicated.

The Lens Model suggests how the aggregated judgment cues implemented in this investigation act as evaluation predictors by identifying the regions of relevance to which they are most likely to be ascribed. By focusing on cognition in the judgment plane, the impact of complex multi-dimensional user *criteria* can be accounted for without being specifically identified. For example, one user may consider a criterion of currency to be a conditional absolute of an IR system search, and another user considers a criterion of authority, while a third user requires a criterion of clarity. Each of those criteria establishes cues at a higher cognitive level (such as pertinence or utility) for the purpose of judging the retrieved items. For each user these lead to a subsequent evaluation of the item in terms of its relevance to the problem at hand. It is assumed that underlying multi-dimensional criteria serve as the basis for evaluating the outcome(s) of judgment (Einhorn, 1978). However, the judgment, not the criteria for judgment, determines the accuracy of the predicted evaluation according to the Lens Model.

There is also strong support for consideration and investigation of the middle ground, since relevance from a user judgment and evaluative perspective is demonstrably not dichotomous in the relevant/not relevant sense. It appears from these analyses that combinatorial judgments represent a user's state of mind regarding the importance of an item retrieved form an IR system at the point in time that such judgments are made. Without such combinatorial judgments, evaluative determinations restricted to single level dichotomies can only describe importance or value to the user, but cannot predict it.

Acknowledgment and understanding of the middle ground should enable the construction of more human-friendly retrieval systems. Discussion of the nature of this middle ground follows.

CHAPTER 5

EXPOSITIVE DISCOURSE

Introduction

For users interacting with IR systems, retrieved items are in the cognitive domain of possible solutions to an information problem. A cognitive domain consists of objects that an individual treats as functionally equivalent based on those attributes by which those objects are comprehended (Scott, 1969). These attributes may serve both to describe the items and to convey the user's attitude toward those items as an expressed response. That response is manifested as an evaluative grouping of attributes capable of conveying the user's point of view toward the retrieved item under consideration. As such a grouping, relevance in the cognitive domain is an evaluative label that conveys a user's comprehension of the item under investigation. This study has explored the cognitive domain of relevance by examining how users group meaningful properties of retrieved items based on evaluative categories as illustrated in Figure 5.1.

The language of this discussion is framed by the Lens Model and data analysis described. The attributes of items retrieved from an IR system are considered to be first order judgments or levels of relevance that users comprehend as being meaningful for the application of an evaluative label which, in turn, is considered as a second order judgment. Topicality, pertinence, utility, systematic and motivational levels of relevance have been treated as a group of meaningful attributes (levels of relevance) that users are able to group into distinctive evaluative labels (regions of relevance).

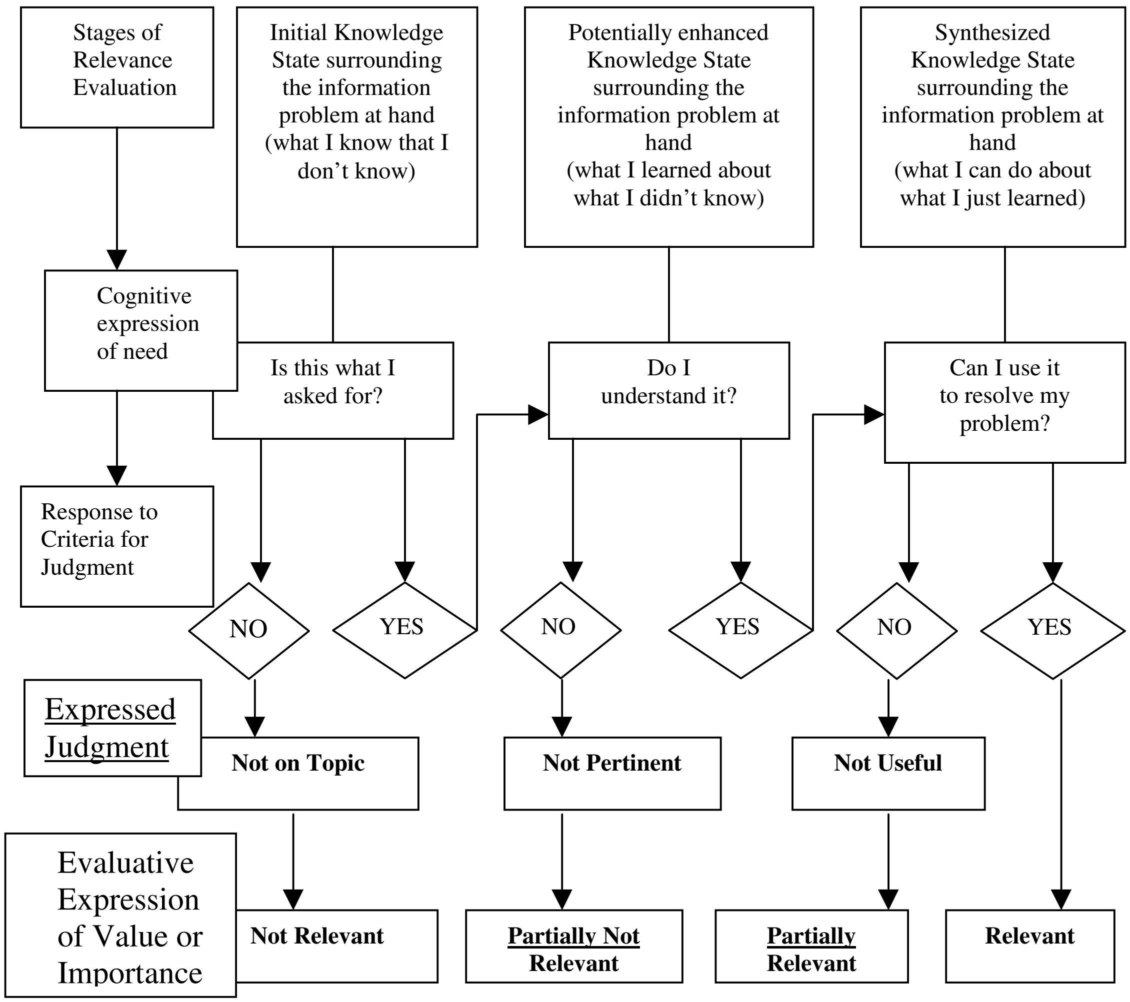

Figure 5.1 Model of End-User Cognition as a Multi-Stage Process of Relevance Evaluation during IR System Interactions

These evaluative labels are the second order judgments or regions of relevance that users assign to their aggregated differentiated first order judgments. Evaluative labels that include relevant, partially relevant, partially not relevant and not relevant designations have been used as the categories of investigation for this study to determine how users heuristically apply these values in relation to their aggregated judgments.

Research Question Findings

The basic question posed by this investigation is whether users of IR systems maintain distinct consistent evaluative heuristics for determining different regions of relevance. The findings imply that judgment consistency across users appears to manifest itself as conjunctive/ disjunctive thresholds that are capable of predicting the region of relevance to which items retrieved are most likely to be allocated. While conjunctive / disjunctive thresholds may be characterized by dichotomous labels at the judgment level, there appears to be a greater range of responses at the evaluation level where simple dichotomous values are aggregated into meaningful regions of relevance.

These findings support recent work suggesting that decision makers attempt to simplify a given decision problem by focusing on a limited subset of attributes and screening out alternatives that fall short on important attributes (Montgomery & Willen, 1999). The results of this study suggest that relevance as a judgmental process leading to evaluative measurement is both a problem solving and decision making exercise involving facilitative cognitive processing. That processing appears to emphasize topicality, pertinence and utility as judgment alternatives capable of defining evaluative measures of relevance. Each alternative alone, however, may fall short of encompassing how the item retrieved is fully comprehended by the user for evaluative purposes.

Relevance as a Decision Making Process

The Lens Model methodology imposed on this investigation implies a relationship between inputs (the set of alternatives implemented by the decision maker) and output (the evaluation of alternatives made by the decision maker). Similarly Carroll and

Johnson (1990) suggest a model of relevance that reflects the form $Y_i = f(X_{ij})$, where Y_i is the evaluative alternative and $f(X_{ij})$ represents a function of the attributes contributing to that evaluative alternative (where the subscript *j* indicates the possibility of many attributes). Weights applied to each of the attributes generally comprise the $f(X)$. A form of such weighting was implied by the instructions given to the users in this investigation. Asking users to identify the levels of relevance most important for making their evaluations triggered a weight. This study, however, did not explore which attribute(s) within an aggregated evaluative response carried the most weight for each user.

The results reported cast a descriptive model of relevance within a decision-making framework by identifying user perceptions of the attributes (levels of relevance) that underlie their preferences during evaluation. Such a model can provide predictions of subsequent evaluations. For example, a user judgment that indicates an item to be not on topic with no other attributes considered is most probably an item that will be evaluated by a user as not relevant. However, a judgment that indicates an item is on topic is most likely to invoke consideration of additional attributes before an evaluative decision is made regarding the item under investigation.

Complexity in the Relevance Domain

Complexity in the Measurement Dimension

The problem of measurement in information retrieval research is not really that of measuring relevance, but of employing relevance as a measure, or more accurately, as a criterion of measurement (Ellis, 1996). Contributing to that problem is the multi-step subjective processing that is required to arrive at a level of cognition that can act as a measurement criterion. Research terminology for defining this type of end-user subjective processing has generally considered measurement in the relevance domain as a single step from document retrieval to evaluation. The results of this investigation point to a more complex process that accommodates the unique aspects of individual user cognitive experience during information retrieval and the underlying multi-stage heuristics that bring those unique user characteristics into an evaluative measurement mode.

Scott (1962) defines cognitive complexity as the number of independent dimensions-worth of concepts an individual brings to bear in describing a particular domain of phenomena. In the field of information science, research surrounding a framework for relevance has tried to narrow in on the independent concepts that individuals bring to bear in describing the relevance domain. What has generally been missing from that research is exploration of the dimensions of those concepts that can be used to frame a theory of relevance.

This study has focused on a set of dimensions that allows users to implement a decision process more closely akin to the conditions encountered in real life search and retrieval situations. Those dimensions include consideration, differentiation and integration, all of which have contributed to the findings in this investigation. Relevance complexity in the measurement domain should account for these dimensions if meaningful assessments are requested from users of IR systems.

Consideration as a Dimension of Complexity in the Relevance Domain

The topicality, pertinence and utility of a retrieved item appear to be important considerations for evaluative purposes by a user. Systematic and motivational considerations do not appear to maintain the same level of importance to users except as supporting levels of relevance as users aggregate judgment cues to formulate evaluative responses of importance or value.

The results of this investigation imply that users tend to implement a *bottom-up* negative to positive approach to evaluation. This is made evident by finding that users are more likely to use only one level of relevance (not on topic) to evaluate an item as not relevant, whereas four to five levels of relevance appear to formulate evaluations in the relevant region. This may help to explain the evaluation process by implying that users may look first at evaluating the item in relation to the query before evaluating the item in relation to the problem. Although not all 'not on topic' items are discarded by users, it may be the first feature considered during the evaluation process. If the item is considered to be so 'off target' that it has no other value, it is relegated to a categorization of not relevant without further consideration.

Consideration in the region of partially not relevant items is also significant. It appears from this investigation that the pertinence (informativeness) of a retrieved item is the next important feature to users in a bottom-up evaluation process. The implication is that users are most likely to use that level of judgment for evaluating items as partially not relevant. A question associated with that finding is 'if the item is judged not pertinent as the single level consideration identified by the user as important to their decision, why is it not evaluated as not relevant?' The implication is that the user has already considered the item to be on topic or close enough to the topic, yet its not pertinent feature moves it one step above complete non-relevance and becomes the most important feature for evaluating the item in that region. Foreign language documents or documents viewed as too technical could easily fall into such a region of relevance.

Consideration in the region of partially relevant items tends to confirm the prior discussion. Both topicality and pertinence appear to be important in this region, however the addition of utility as an important consideration moves the evaluative process into this next region of relevance. Items evaluated by users in this region, however, are highlighted by the feature that users perceive to be lacking, which as the results have shown are most likely to be a judgment of not pertinent or not useful. A significant contribution to modeling partially relevant items also came from items judged systematic, topical and pertinent. The implications from this finding are twofold. One is that the use of systematic as a consideration represented items that were in forms or formats that were not exactly what the user was looking for, yet a negative judgment for that feature would not have been reasonable to the users. The second implication is that users may have been somewhat unsure of the usefulness of the item even though the other three features were evident, thus a delegation to the partially relevant category as the evaluative response.

From this bottom up perspective on consideration in the relevance domain, it appears that user judgments of items evaluated as relevant consist of the major cues of topicality, pertinence and utility, along with other supporting cues that users consider as features for moving their current information problem towards resolution. While systematic and motivational considerations were introduced as independent variables in this study, those considerations could also be represented as conceptual sub-categories of usefulness. If the item is in the right form or format, it can be considered useful. If an item causes action by the user in the motivational sense, the item can be considered useful. Similar implications surrounding systematic and motivational cues in the negative sense could yield not useful judgments.

This conceptualization of hierarchical multileveled extraction of evaluative features from the cognitive realm of users is supported by the work of Kreitler and Kreitler (1976) as a cognitive amalgamation leading to an active synthesis of an evaluative response. Support for this *bottom-up* heuristic is also represented by the concept of intra-alternative search (Svenson, 1979) where a decision-maker investigates the aspects of one attribute before going on to the next attribute. It appears from the results obtained herein that users (as decision-makers) are most likely to consider topicality, then comprehension (pertinence), and then usefulness in their attempts to evaluate items retrieved from an IR system.

Differentiation as a Dimension of Complexity in the Relevance Domain

Differentiated relevance is implied when task, problem, situation and retrieved item are cognitively merged for the purpose of judgment and evaluation (Ingwersen, 1996). While consideration versus non-consideration act to separate meaningful attributes for judgment purposes, differentiated aspects of those attributes also appear to play a significant role in leading to cognitive evaluation. A visual depiction of these relationships are shown in Figure 5.2.

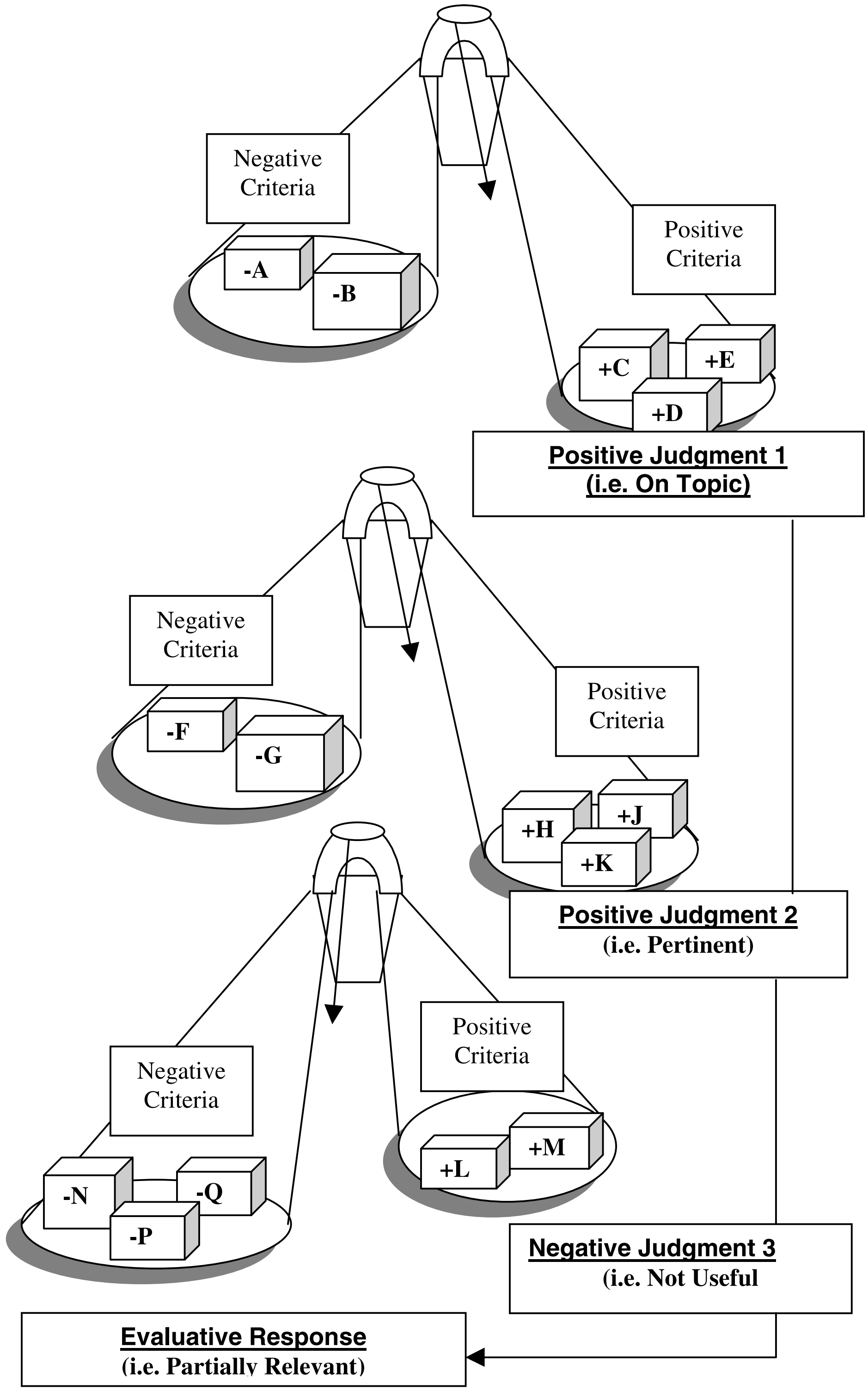

Figure 5.2 Dichotomous Approaches to Differentiation and Integration Leading to Non-Dichotomous Evaluation

Although situational effects may impact the decision process, it seems likely that variables with positive or negative framing would combine as individual values that influence the importance given to different cues (Ebenbach & Moore, 2000). If positive versus negative differentiation among attributes is not taken into account, theoretical perspectives surrounding relevance could remain limited at all levels of description, explanation and predictability.

These positive/negative aspects of user judgment attributes also serve to define a more robust framework of relevance by creating defining terminology for middle regions of relevance that invoke dichotomous perspectives without being limited by them. When it comes to salient beliefs (user judgments), it is quite plausible to assume that salient attributes are particularly high in value, either positively or negatively (Sjoberg, 1999).

Most contemporary decision theories do not consider degree of differentiation which represents how important different attributes are to people at a given point in time (Svenson, 1999). However, Taube (1965) in his discussion of the pseudo-mathematics of relevance was one of the first to recognize that in the context of a searcher's judgment, relevance is a psychological quality that describes the positive or negative relationships of meaning between a document and the problem at hand. The importance of that differentiation has been re-established by this investigation.

<u>Integration as a Dimension of Complexity in the Relevance Domain</u>

Since cognitive components are virtually unlimited, a search for a finite number of contributing variables to model evaluations in the relevance domain requires analyses that account for $nD^v - 1$ combinations of cognitive input. These judgment cues (nD) equal the number of dimensions of an individual variable being investigated to the power V that represents the number of variables being investigated. In this study $3^5 - 1$ possible judgments were available to users for structuring first order judgments used to evaluate retrieved items. Of these 242 possible combinations (involving three dimensions including positive consideration, negative consideration and non-consideration on five variables including systematic, topical, pertinence, utility and motivational levels of relevance) users implemented 103 different combinations of judgment cues to evaluate 1432 items in four regions of relevance (not relevant, partially not relevant, partially relevant and relevant). Of the 103 different judgment cue combinations, 45 judgment combinations were used to describe more than one region of relevance.

These results provide further evidence of the difficulty of addressing finite representations of relevance when there are virtually no limits to the cognitive content used to define those representations. The search, as implied by Newell (1990), is not a search for finite variables, but a search for finite combinations of variables that define the problem spaces that users construct surrounding their information needs. Reducing relevance to only on/off or yes/no analyses on the basis of topical judgments subjugates user cognition in the relevance domain to terms that are insufficient for effective user evaluative processing.

It is difficult to specify criteria to test methods for mathematically aggregating evaluative judgments. In part, this is because value is inherently a subjective judgment, and in part it is because of the ill-structuredness, complexity, and high dimensionality of the situations for which aggregated judgments are desirable (Ferrell, 1985). For these reasons, the approach to this investigation was to look at user aggregated judgments as the threshold for relevance evaluations; and it appears from the results that users do implement aggregating principles to define their structured evaluative responses.

The Structure of Relevance Evaluations

Relevance evaluations appear to be predicated on conjunctive and disjunctive thresholds of judgment. Judgments can be characterized as aggregated features, attributes or dimensions of items retrieved from an IR system that cognitively map the importance or value of that item to the user in relation to the current information problem at hand. Differentiated aspects on these different attributes generate the decision-maker’s perceived pattern of avoidance or attraction to the available decision alternatives. For example, an examination of one attribute taken from Figure 5.2 shows that the weighted positive versus negative values of underlying criteria contribute to an integrated judgment decision as described by Figure 5.3.

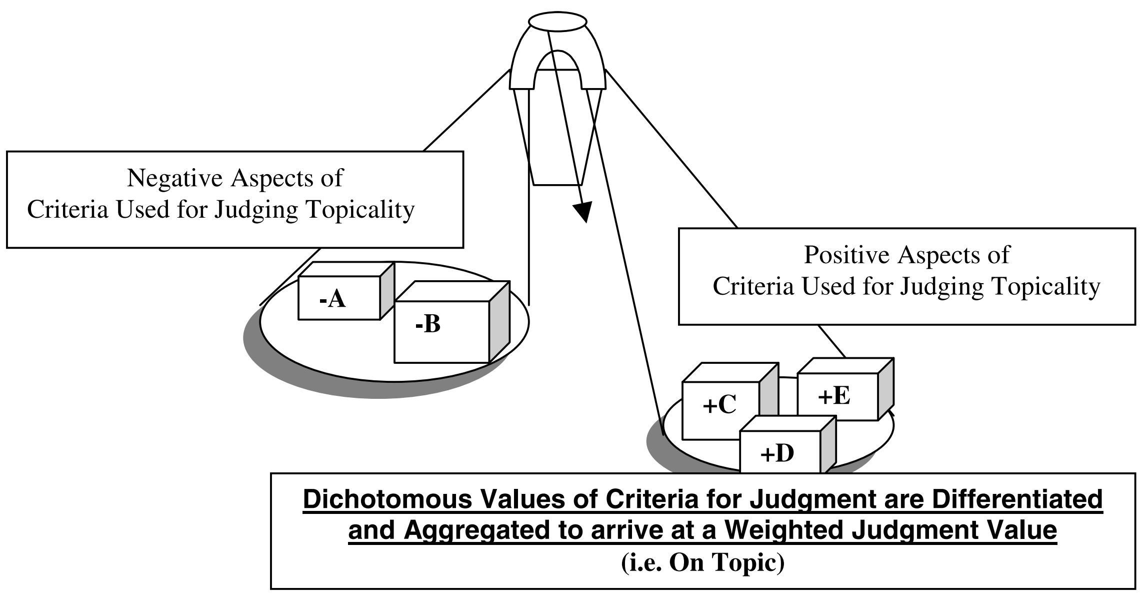

Figure 5.3 Judgment Decisions as Differentiated Aggregations of Underlying Criteria

In the research arena, decision-making has been described as a process where successive applications of decision rules constitute the cognitive sub-processes reflecting the search for cues necessary for a decision (Svenson, 1979). The dominance rule states that one alternative will be chosen over another if it is better on at least one attribute and not worse on any other attributes. Based on this decision rule, it appears from the results of this investigation that before a user determines if an item can resolve his/her problem, he/she first looks to see if the item is about the problem (topicality) before any other attribute of the item is considered. That thinking is supported by the anchoring and adjustment heuristic (Slovic, 1972; Tversky & Kahneman, 1974; Wallsten, 1978) stating that a judgment is initially made on the basis of a particular most salient dimension of the stimulus, and is then adjusted as additional dimensions are considered.

The conjunctive decision rule requires that before an evaluative choice is made, a set of attributes associated with that alternative must be evident or the alternative is dropped from consideration. The findings in this study point to the use of this rule by users as they consider which region or degree of relevance the retrieved item under investigation represents.

The disjunctive decision rule requires that before an evaluative choice is made, there is at least one aspect from a set of attributes greater than all the others that prompts the decision on that alternative. These results confirm that users apply the disjunctive rule as a heuristic for evaluating items retrieved from an IR system by treating the negative aspect of a judgment cue as a cognitive prompt for choosing some region of relevance for that item. For example, if an item is judged not on topic it is most likely, but not always, considered not relevant. If the item is on topic, but not understandable (pertinent), it is moved up a notch on the relevance measurement scale. If the item is on topic, is understandable, but not useful in resolving the current problem it may only be partially relevant. If the item is on topic, not entirely understandable, but considered useful, it may also be delegated to the partially relevant evaluative category. Only when there is no disjunction in the user's set of applicable judgment cues will the evaluative response emerge in the "relevant" category as a measure of the item's importance to the information problem at hand.

While these decision rules may point toward an explanation of user cognition during IR evaluations, they do not reflect the only explanations afforded by prior research in problem solving and decision making disciplines. Payne (1976) suggests that individuals sometimes combine strategies, typically with an initial phase where poor alternatives are eliminated and a second phase where the remaining alternatives are examined in more detail. The minimized consideration of systematic and motivational judgment cues identified in this investigation lends support to that idea. Elimination of poor alternatives was first suggested by Tversky (1972) and this EBA heuristic (elimination by aspects) is also supported by the evidence provided in this study. Although researcher-imposed judgment cues were provided, users appeared to eliminate those cues from consideration when they did not meet some subjective value for making the required evaluative decision.

Aschenbrenner, Bockenholt, Albert and Schmalhofer (1986) suggest decision making is the result of the majority of confirming dimensions (MCD) heuristic. They assert that attribute differences are processed sequentially, with the summed differences accumulating until the advantage of one option over the other exceeds some criterion value. The bottom-up approach to relevance evaluation that is supported by this investigation would appear to confirm that users may implement such a MCD heuristic to make relevance evaluations particularly in the region of high relevance where a positive summation of multiple judgment cues appear to contribute to the evaluative decision.

Screening (Beach, 1993) may also be an important heuristic in relevance evaluation. As users reflect on a particular option (levels of relevance), they may bring to bear their own goals, values and beliefs applicable to the decision problem resulting in a decision that is compatible with their own standards on each cue presented. Topicality has remained the major focus in relevance evaluations because it can be defined objectively. Pertinence and utility, however, are difficult to evaluate without bringing subjective considerations into the analysis.

Attempting to define an overarching structure for relevance evaluation by users of IR systems, makes evident that the problem solving and decision making characteristics of this cognitive activity embrace a large number of rules and/or heuristics. Even after experiencing considerable vacillation (Janis, 1977), a decision maker reaches a point of feeling confident that the best choice has been made. It would appear from this investigation that the conjunctive/disjunctive judgment thresholds established by users tend to structure a best choice for evaluating regions of relevance even during periods of possible vacillation.

If all of these individual heuristics are implemented by users of IR systems to construct meaning during the search for a reasonable solution to their information problems, then the goal of the decision process is not just to fulfill one or several rules (or heuristics), but to apply several approaches in order to find an alternative that is sufficiently superior to any others (Svenson, 1999). As aptly stated by Dretske (1981), the 'optimal choice' is the one with the *best worst state*.

User Attitudes Toward Information

Attempting to view and understand human cognition has consistently posed problems for empirical research. No single individual, by nature or nurture, can be assumed to behave in a consistently predictive manner. User predictions of relevance are no exception. However, if an understanding of user attitudes toward information is established, that predictability might be significantly enhanced.

If attitude is considered an internal state that affects an individual's choice of personal action toward some object (Gagne, 1985), then users can be expected to have attitudes toward information-laden objects retrieved from IR systems. While the prior theoretical perspectives on problem solving and decision making may help to advance a better understanding of user cognition during relevance evaluations, additional theoretical frameworks can also contribute supporting evidence for a better understanding of user evaluative attitudes. A basic reason for incorporating a discussion of attitudes in this study is that the essential characteristic of an attitude is its evaluation component viewed generally as a summary statistic stored in memory (Pratkanis, 1989). As such it serves mental classificatory purposes that influence reasoning, decision-making, interpretation, and inference, all of which can be considered cognitive components of user relevance.

Relevance as an End-User Attitude Toward Information

Relevance has generally been used as a measure of retrieval effectiveness, yet it appears from an end-user point of view that relevance is a measure of (attitude toward) problem resolution potential. The retrieved item itself has no intrinsic value. The inferred relationship that connects a retrieved item to a user's information problem is an attitudinal response on the part of a user. Thurstone (1931) was among the first to indicate that a person's attitude is a distribution of values rather than a single point on a measurement continuum. From the results of this study, it could be postulated from the Lens Model representative design that when users are experimentally 'forced' to identify a measure of relevance on a scale by designating a single point (or single categorical value), they do so by aggregating their attitudes surrounding the available (from experimental imposition or memory) attributes to provide a single point best estimate of evaluation. This interdependence of attitude theory and measurement adds an additional framework for developing a more expositive view of user evaluations during IR interactions.

The cognitive view of attitude requires four concerns that include (1) identification of thought content, (2) the subjective properties of that content, (3) how those properties are aggregated, and (4) an identification of the response domain (Ostrom, 1989). The worksheet and instructions provided users in this study were designed to stimulate (1) and (2) in a pre-established investigative direction, uncover the user heuristics surrounding (3), and (4) impose a response domain (regions of relevance) consistent with prior work in the field. From that perspective it could be argued that relevance as a conceptual framework may benefit from prior attitudinal research, with particular regard to the fact that attitudes are covert acts that project meaning to a position on some dimension of judgment (McGuire, 1989). Relevance evaluations, in general, are such covert acts on the part of IR system users.

Another issue derived from this investigation is the underlying assumption that user cognition leading to relevance evaluations is a heuristic exercise. That is, users appear to consistently use some type(s) of guiding principles for uncovering the value of an item retrieved from an IR system. That may, however, not be entirely true. Howell and Burnett (1978), demonstrated that processing in the realm of uncertainty judgments may result from combinations of four classes of cognition that include belief in specific required characteristics (i.e. rules of thumb; pre-established criteria for judgment), stored historical data (existing knowledge surrounding the problem at hand), heuristics (strategies for removing uncertainty), and systematic biases (characteristics of the situation). While an attempt has been made to uncover user heuristic approaches to relevance evaluation in this study, it is recognized that user attitudes encompassing each of the above classes of cognition may confound modeling of how users evaluate items retrieved from an IR system.

Another question arises from the Lens Model methodology prescribed for this study in which it was assumed that judgments could represent user attitudes. Can the first order judgments based on the positive or negative characteristics of the various levels of relevance (topicality, pertinence, utility, systematic, motivational) be considered as user attitudes? Or, do user evaluations of regions of relevance (not relevant, partially not relevant, partially relevant, relevant) represent user attitudes? According to Pratkanis (1989), both can be considered as user attitudes about the retrieved items since an attitude is represented by a simple evaluative cue (differentiation) which serves a heuristic function (aggregation) leading to a schematic function (evaluation). In this sense, the attitude is the heuristic for resolving the problem at hand, which can be posed as, 'Do I need this retrieved item or not?' More aptly, an attitude provides an aid for 'sizing up' objects in the environment (Smith, Bruner & White, 1957), which is exactly what users are doing when confronted with what may be largely database detritus. The unknown user attitudes in this study are the underlying cues (criteria) that users differentiated and aggregated to get to their first order judgments. For example, users were not asked to identify why an item was/was not pertinent or why an item was/was not useful.

The dichotomous nature of relevance at initial levels can also benefit from the inclusion of attitudinal perspectives, since evaluative bipolarity has been the focus of most past attitude research (Ostrom, 1989). On an attitude continuum people tend to lean in one direction (positive) or the other (negative) based on the relationship being considered. In this case, the relationship is end-user information need to end-user information problem at hand. In addition, Sherif, Sherif and Nebergall (1965) claimed that attitude should be represented as regions on such a continuum. These prior studies further the argument that user cognitive evaluations of retrieved items are judged by differentiation on their positive and negative aspects and that those evaluations can be grouped into regions of relevance as an attitude toward the potential resolution of their information problem.

Some attitude research has addressed cognitive differentiation and integration in a manner similar to that used in this study. Ajzen (1989) in a discussion of differentiation versus evaluative consistency indicates that there is a tendency for evaluative consistency to be inversely related to differentiation. That is, the greater the number of dimensions evaluated the less evaluative consistency will be found. This tendency combined with similar results from most relevance research leads to the conclusion that user relevance evaluation is a complex multi-dimensional matter that is difficult to address in simple terms. However, consistent with Ajzen, this study shows that by limiting the number of dimensions and controlling the aspects of differentiation, evaluative consistency can be achieved in terms of conjunctive and disjunctive thresholds of relevance.

Ajzen also asserts that there is a tendency for evaluations at the extremes to decrease with differentiation. In terms of positive conjunction, negative conjunction, and disjunction, the findings in this study support Ajzen’s assertion. That is, there is little differentiation at the extremes as represented by positive conjunctive attitudes about retrieved items at the relevant extreme or negative conjunctive attitudes at the not relevant extreme. However, once there is differentiation in the disjunctive sense (mixed positive and negative attitudes about retrieved items) evaluations no longer appear at the scalar extremes.

Attitude research also lends credence to the middle range of partial relevance. Ostrom and Upshaw (1968) in their discussion of multiple anchors assert that an individual's attitude is not only anchored at the extremes, but also at mid-scale. The mid-scale anchor being formed through the selection of content midway between the two perspective end anchors. In this investigation, user attitudes toward the items being evaluated were anchored at the not relevant end of the scale by the conjunction of negative judgment cues and anchored at the relevant end of the scale by the conjunction of positive judgment cues. In the middle regions of partially relevant and partially not relevant evaluations, users were anchored by the selection of disjunctive (positive and negative) judgment cues. Ostrom and Upshaw also indicate that mid-scale attitude anchors have independent determinants. Evidence from this study supports this assertion. Negative aspects of pertinence appear to be more of an anchor at the lower end (low middle) of the relevance scale, whereas negative utility judgments appear to be more of an anchor at the higher portions (high middle) of the scale.

Because a dimension (in this case relevance) has a middle as well as two ends, it seems reasonable to suggest that all information located along that dimension could be considered as fitting it (Eagley & Chaiken, 1993). As such, information portrayed to users in the middle range of relevance is just as important as the information portrayed at the extremes. Exactly what the predictive nature of these middle values represents, however, is still subject to additional investigation beyond the scope of this study. Also, these findings do not necessarily demonstrate that users can differentiate relevance evaluations into only four regions of relevance using only five levels of relevance.

In addition, and consistent with the findings of this study that point to the important nature of positively and negatively differentiated judgments as thresholds to relevance evaluations, van der Pligt and Eiser (1984) provide research evidence to support that negatively labeled information is more influential than positive information in the formation of an overall evaluation. If that is the case, then it would appear that while users seek what they need, they are continuously filtering out what are perceived as negatively evaluated items as judiciously and economically as possible, first by topic, then by informativeness, and then by usefulness.

Application to Relevance Theory

A theory is a set of interrelated constructs (concepts), definitions, and propositions that presents a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena (Kerlinger, 1967). Relevance is such a set of interrelated constructs, definitions, and propositions that represent a systematic view of user evaluative behavior during information search and retrieval activities.

Since the emergence of relevance as a key construct (concept) in information science, a large number of variables has been identified providing evidence that relevance is a phenomena that can be viewed in theoretical terms. The findings of this study point to heretofore undefined relationships among variables. These relationships enable a more systematic view of relevance capable of greater descriptive, explanatory and predictive power.

First, user evaluations of items retrieved from an IR system appear to be a multi-stage process. Each of those stages encompasses variables that structure cognitive relationships (levels of relevance related to regions of relevance) that users implement to make decisions in order to resolve their information problems. Constructs within these various stages, such as criteria for making judgments, dimensions of judgment, the judgments themselves, as well as the evaluative responses, each require their own definitions. A systematic view of user behavior that envisions relevance as a measure of retrieval effectiveness should not treat these variables as synonymous.

Second, relevance as an evaluative construct requires interdisciplinary perspectives derived from problem solving, decision-making and attitude theoretic approaches to fully describe, explain, and predict user behavior during interactions with IR systems. While this study has made no attempt to be exhaustive in that regard, it does provide evidence that relevance theory surrounding information retrieval and evaluation requires additional clarifications that may emanate from these interdisciplinary frameworks.

Each of these issues has been incorporated into the operationalization and analysis of the original research question in order to direct attention to some of the underlying difficulties associated with relevance research and relevance as a theoretical evaluative framework in information science.

Summary

This investigation has been a response to the emergence of concerns regarding user cognition during interactions with IR systems. These concerns have found expression in recent years centered on user judgment and evaluative patterns that can more closely match user needs to items retrieved from IR systems. This study has been in direct response to those that have called for such investigations with statements such as:

Future work in user modeling should concentrate on the issues of evaluation and the resolution of competing inferences (Daniels, 1986);

Evaluation must be based on something more than how well the user can formulate a query. Evaluation must be multi-dimensional giving the user a chance to be as discriminating as possible (Tague & Schulz, 1989);

Relevance is constructed in a user's mind, a clue to one user is not necessarily a clue to another user (Harter, 1992); and,

Considering that relevance is cognitive, subjective, situational, complex, multi-dimensional, dynamic, systematic, observable, and measurable at a single point in time, and that users evaluating information within the context of a current information need will base their evaluations on factors beyond topical appropriateness, understanding a core set of relevance concepts can benefit both basic and applied research (Barry & Schamber, 1998).

In response to the above, it appears from the results of this study that there are certain cognitive characteristics that identify how users move through the process of evaluating items retrieved from an IR system for the potential resolution of their information problems that include:

(1) A multi-stage evaluative process that begins with establishing criteria for making relevance judgments, making aggregated assessments at a higher cognitive level of judgment (levels of relevance), and then aggregating those higher level judgments into single evaluative responses of overall relevance (regions of relevance) to the problem at hand;

(2) Levels of relevance possess bipolar characteristics (positive and negative) which allow users to discriminate and differentiate how they apply to their problem at hand;

(3) Topicality, pertinence and utility appear to be major levels of relevance that users can cognitively differentiate dichotomously for making value judgments on a relevance continuum;

(4) Users aggregate those levels of relevance to establish a single valued label (regions of relevance) that can be ascribed to the importance of a retrieved item in relation to their problem at hand;

(5) Those single valued labels (regions of relevance) exist as a range of values on a mental continuum from which users can determine what steps to take next in their information seeking process in relation to the individual item(s) retrieved; and,

(6) Regions of relevance can be identified on an evaluative continuum by the conjunctive and disjunctive nature of the aggregated judgments used to make a single valued summation of importance to the problem at hand, where conjunction is prominent at the continuum extremes and disjunction is prominent in the mid-scale regions.

Relevance is a key concept and evaluative measure in the realm of user interactions with IR systems. The results of this study indicate relevance has characteristics of phenomena investigated in a wide variety of disciplines under rubrics such as problem solving, decision-making and attitude structures. All of these draw from issues of cognitive complexity surrounding terms that include dimensionality, differentiation and integration in a multi-attribute space.

CHAPTER 6

CONCLUDING SYNTHESIS

Introduction

The intent of this investigation was to advance research on relevance as an evaluative framework from an end-user cognitive perspective. The methodology, data analysis and discussion were founded on evidence that human evaluations of objects in their environment consist of multi-stage processes that engender heuristic approaches that lead to evaluative decisions. Items retrieved from IR systems and evaluated in relation to an information problem at hand fall into this realm of multi-stage cognitive processing. To more fully understand the end-user during such IR system interactions, these multiple stages require more robust identification, description and explanation. This study has taken steps toward a more effective IR system by suggesting a framework and demonstrating its efficacy.

Assumptions Affirmed

Two major assumptions girded the framework of this investigation. First, relevance as an evaluative concept in information retrieval is not a single dichotomous decision process. Second, users can make evaluations of relevance based on at least five broad-based characteristics of relevance termed systematic, topicality, pertinence, utility, and motivational considerations. The analyses of the data collected for this study affirmed both of these underlying assumptions.

Whether relevance exists categorically or on a continuum in the minds of users is not an issue of this investigation. It appears, however, that users have the capability of categorizing items retrieved from IR systems into more than simply relevant or not relevant. As originally suggested by Katter (1968), users are capable of many categorical segmentations of relevance (absolutely none, near minimum, weak, noticeable, very noticeable, strong, very strong, near maximum, and maximum). However, on a user-by-user basis the conceptual meanings of those various categories become blurred and hence lose their value as key indicators of measurable differentiated relevance. It is clear that user cognition incorporates a range of values, the complete nature of which is yet to be determined.

Attributions of relevance lend structure to investigative studies of user interpretations of retrieval value as evaluative regions of relevance. The attributes chosen for this investigation, derived from a research consensus that encompasses retrieval evaluation from topical, pertinence, utility, systematic and motivational considerations, have the capability of providing users with the conceptual frameworks necessary to make evaluative decisions. However, these attributes (levels) of relevance are by no means to be considered the only features of retrieved documents capable of providing users with the judgment characteristics for evaluative decisions of relevance.

Major Findings

Conjunctive/Disjunctive Thresholds of Evaluation

Prior discussion has suggested that end-users differentiate and aggregate features of retrieved items to express labels of importance associated with the value of those items. Conjunctive thresholds of evaluation are represented by aggregations of features with a least common denominator, which in this study is represented by the positive or negative nature of the underlying features. Disjunctive thresholds of evaluation are represented by aggregations of features that have mixed positive and negative values.

A major finding of this investigation is that, given a set of parameters for judging the importance of items retrieved from an IR system, aggregated conjunctive and disjunctive characterizations of those judgments act as thresholds for modeling a cognitive spectrum of values in relation to an end-user information problem at hand. Relevance, as an evaluative construct of retrieved item importance, can be measured on that spectrum of values.

An additional finding is that the nature of that spectrum of values consists of at least three regions. The first is a region of negative conjunction, where end-users generally evaluate items as not relevant due to an aggregation of negative aspects. The second is a region of positive conjunction, where end-users generally evaluate items as relevant due to an aggregation of positive aspects. The third is a region of relevance where end-users generally evaluate items as having some degree of partial relevance due to an aggregation of disjunctive aspects.

Based on prior studies that incorporate different types of relevance scales as evaluative measures of retrieval effectiveness, this study leads to a synthesis of concepts of user cognition that supports multiple judgment combinations for describing the importance of items retrieved from an IR system. Although additional research is required, it would appear that these various conjunctive and disjunctive judgment combinations act as cognitive points of discrimination for evaluative labeling without concern for the number of measurement points on the scale. Negative conjunctive judgments define one end of the scale (not relevant), positive conjunctive judgments define the other end of the scale (relevant), and disjunctive judgments span the middle regions of the scale no matter how many points or categories are presented. If scale values are forced into a strictly dichotomous scale (not relevant / relevant), then disjunctive judgments will confound the measurement value of the measurement instrument since they will create a bias at one end of the scale or the other.

Topicality, Pertinence and Utility

Analysis of the data collected for this investigation affirm that topicality is a major feature of items retrieved from an IR system for the purpose of resolving an information problem at hand. As a single feature of such items, however, topicality appears to be more important to users for de-selecting items than for selecting them. That is, unless some other feature(s) of the retrieved item are considered, topicality by itself will not determine if an item is relevant; it is only an indication of potential relevance.

Pertinence or comprehension of the item retrieved and the information it conveys to the end-user appears to be the next major feature of importance to end-users. Without this cognitive connection of retrieved item to problem at hand, no further evaluation can take place. As a single feature of items retrieved from an IR system, pertinence reveals (through understanding) or hides (through a lack of understanding) what end-users already know about what they already know, what end-users thought they didn't know, or what end-users didn't know they didn't know. Judgments that incorporate pertinence as a cognitive attribute of relevance evaluations are for the most part subjective in nature.

The usefulness or utility of an item retrieved from an IR system in relation to an information problem at hand is an indication of the item's ability to prompt further action. As a single feature of items retrieved from an IR system, utility is an attribute related to the action a user will take once they have an item that represents that value in time. Topicality and pertinence speak to the user's best guess here and now of the likelihood of the item resolving the information problem at hand, while utility speaks to the actual resolution of the problem at hand sometime in the future. The results of this investigation suggest that until issues of topicality and pertinence are determined by the end-user utility is difficult to evaluate or predict. Indeed, utility may be said to be an *a posteriori* attribute of the evaluation process. Utility was only used as a single attribute judgment 43 times (3%) of the 1432 items judged.

It is important, though, to expand the discussion so that the utility issue is not misconstrued. For example, items retrieved for use in developing this very study were collected based on their topical importance to the issue under investigation and their meaning to the author for engaging the problem at hand. The utility of each item, however, was first evaluated during collection, again during the data analysis, and finally during the discussion of the findings. Several iterations of utility judgments rendered the original items to be either relevant, partially relevant or not relevant depending on the point in time that those judgments were made; while at the same time, relevance based on topicality and pertinence did not change significantly.

In addition, items retrieved from an IR system that are not on the topic requested and/or not pertinent to the problem at hand can still be useful for some other purpose. This discussion and the results of this investigation have centered strictly on the end-users information problem at hand as conveyed by the instructions to the participants in this study and have not addressed or analyzed utility from that perspective.

Areas for Further Investigation

The preliminary work leading to this study and the results of this investigation point to user relevance evaluations as wide ranging judgments that are cognitively differentiated and aggregated to arrive at single value designations of importance. These single designations of importance can be modeled into regions of relevance that separate these various judgments into a meaningful spectrum of values which create a relevance continuum by bridging the gap between non-relevance and relevance for the end-user. The middle region(s) of relevance created by this values-based continuum of judgments consists, for the most part, of a disjunctive cluster of retrieved item features. Although beyond the scope of this study, the weighted value(s) of those aggregated disjunctive features require further investigation in order to discern a full understanding of the user relevance spectrum.

Metaphorically, the color spectrum of relevance as derived from this study points to a red (not relevant) to violet (relevant) continuum that moves from topicality to pertinence to utility as end-users seek to gain value from items retrieved from an IR system. To further clarify this view of user evaluative behavior during IR interactions as envisioned in Figure 6.1, additional validation is required.

Relevance Spectrum (Levels of Relevance)

<table>
<tr><td colspan="3">???</td><td colspan="4">Topicality</td><td colspan="5">Pertinence</td><td colspan="4">Utility</td><td colspan="4">???</td></tr>
<tr><td colspan="20">Shades of underlying features (criteria) leading to judgment</td></tr>
<tr><td>?</td><td>?</td><td>?</td><td>?</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>?</td><td>?</td><td>?</td><td></td></tr>
</table>

<u>Figure 6.1 Filling in values of a relevance spectrum</u>

The total construct of mediating variables that constitute a cognitive relevance spectrum may vary from user to user; however, it appears from the results of this investigation that a TPU (topicality, pertinence, utility) approach represents a replicable framework that could further an understanding of user interactions with IR systems.

Definitional Concerns

Terminology associated with relevance research has generally been inconsistent, incongruous and sometimes confounding in presenting a solid foundation for succeeding investigations. Analyzing end-user evaluative processing as an underlying framework for understanding relevance in a theoretical context can provide a means for greater definitional clarity. The following definitions are derived from reflection on the conduct and analyses of this study:

- *Criterion* (for judgment): A single level concept that describes a feature of an object (retrieved item) for evaluative purposes.
- *Judgment*: An aggregation of criteria that describes an evaluative framework for selecting and de-selecting an item related to an information problem at hand.

- *Topicality*: A criterion found both in an end-user query and in a general level content descriptor field in a database record. It also acts as a cognitive mechanism for separating items retrieved from an IR system based on perceived and expressible unimportance versus potential importance to an information problem at hand.
- *Pertinence*: A judgment (based on underlying criteria) for separating items retrieved from an IR system based on the cognitive comprehension of meaning the items convey to the end-user in relation to the information problem at hand.
- *Utility*: The 'actionability' the item enables for the end-user in relation to the information problem at hand. Utility implies action in the future, either immediate or pending. Thus, utility as a judgment implies that a next step is, or is not, possible now that the item has been retrieved and so judged.
- *Relevance*: A cognitive construct of importance for an end-user made up of underlying criteria that can be aggregated for judgment purposes on an overarching set of imposed values.
- *Relevance evaluation*: An aggregation of judgments that describe retrieved item importance in a single term (or single point) within a self imposed or externally imposed range of values.
- *Partial relevance*: Any disjunctive aggregation of judgments surrounding the evaluation of an item retrieved from an IR system. A threshold of evaluative certainty within the realm of information problem uncertainty.

Implications for IR Systems Development

Topicality appears to be of value to an end-user because it represents a cognitive heuristic for merging an information problem with IR system design capabilities by matching a requested topic. Topicality, however, while acting as a cue to functionality does not necessarily imply functionality. If a topical match does not occur, the end-user is most likely to ignore or discard the item retrieved. The results of this investigation, however, point to a multi-step cognitive process implying that positive topicality is only a first step, not an end, to end-user cognitive evaluation of retrieved items. If an item is not discarded because it is on the topic requested, the end-user leaves the realm of topicality to determine whether the retrieved item, indeed, possesses the functionality required to solve the information problem at hand as pictured in Figure 6.2.

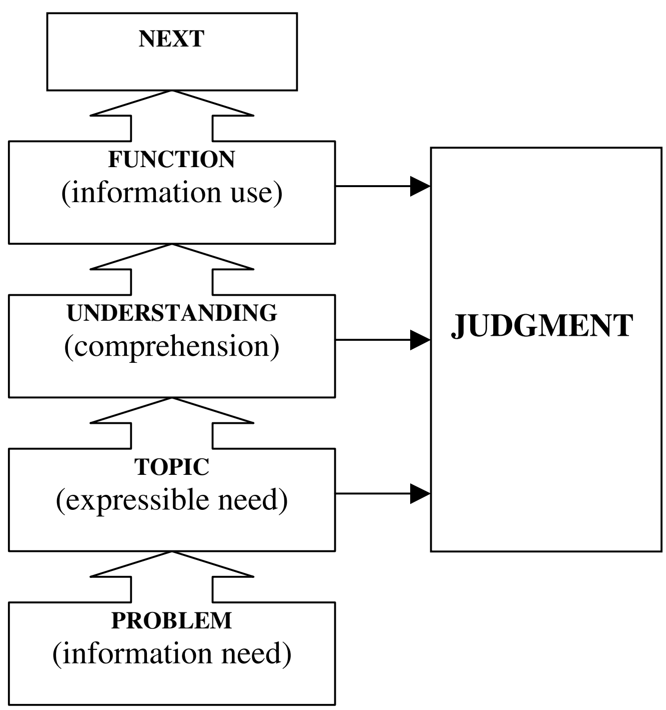

Figure 6.2 Moving from topicality to functionality during the evaluative process

Or, more succinctly, as derived from Wilson (1977), the end-user reorients toward the functional rather than the topical by explicitly recognizing the primacy of the need by cognitively bringing the state of knowledge to the point of use.

To date, IR system design and development approaches have generally attempted to refine (and thus constrain) access to the realm of topicality. Yet, for the most part, they have failed to move toward design criteria that acknowledge the succeeding steps of end-user cognitive evaluative processing. With the exponential growth of information in general and technologies for greater storage and quicker retrieval capabilities, recent calls for more effective indexing of information through the use of applicable metadata schemes have yielded a significant agenda for continuing research (Weibel, 1995; Bearman, 1997). If end-users require information that is not only on topic, but also is comprehensible and useful in relation to their base of knowledge and their information needs, then metadata schemes and their associated retrieval algorithms should include approaches to these realms (pertinence and utility) of user evaluative processing.

To What End…?

The immediate implications for IR system design include the development of meaningful taxonomies of user comprehension (pertinence) and utility (usefulness). With such taxonomies, the implementation of metadata schemes that account for these cognitive realms of evaluation could enhance retrieval effectiveness from both end-user and IR system design perspectives as outlined in Figure 6.3. Recall and precision measures based on a three-value criterion (topic, meaning, use) would more adequately merge IR system evaluations of effectiveness with end-user perceptions of relevance.

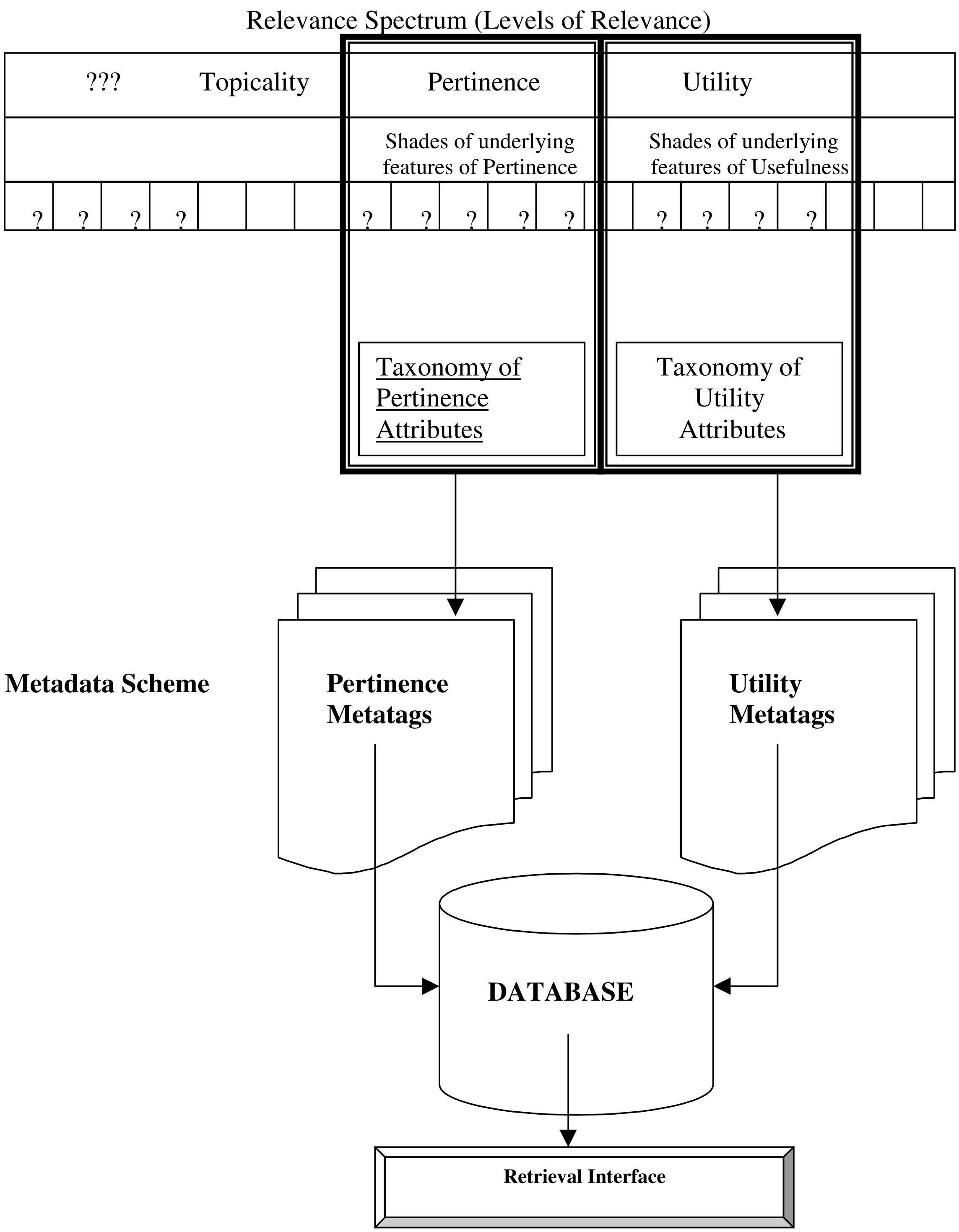
Relevance Spectrum (Levels of Relevance)
???
Topicality
Pertinence
Utility
Shades of underlying features of Pertinence
Shades of underlying features of Usefulness
? ? ? ?
? ? ? ? ?
? ? ? ?
Taxonomy of Pertinence Attributes
Taxonomy of Utility Attributes
Metadata Scheme
Pertinence Metatags
Utility Metatags
DATABASE
Retrieval Interface

Figure 6.3 Implications for design improvement with enhanced taxonomies

The prolific spread of technological developments in the storage, access and retrieval of all types of information has contributed to a need for all individuals, child to adult, to have a basic understanding of information search and retrieval techniques. Continuing developments in Internet, intranet and extranet technologies will further enhance this need. This same level of envisioned development supports IR systems that could be sensitive and hospitable to human search and evaluation modes. While individual approaches to such activities may vary by situation, by knowledge, and by experience, there appear to be underlying heuristics that act as points of cognitive convergence for most end-users during the evaluative processing of information.

One of the goals for research in information science should be to uncover the cognitive heuristic processes that can be replicated in IR systems. To date, algorithmic match of query topic to database items represents a necessary but insufficient solution for maximizing search and retrieval effectiveness for an end-user. Although weighting schemes demonstrate understanding of more than dichotomous ranking, the criteria on which the weighting has been accomplished have not reflected the variety of user cognitive constructs necessary for satisfactory problem resolution. Lens Model representative design as implemented in this study provides a framework for further investigation that can shape how end-users make decisions during IR system interactions. Figure 6.4 links the Lens Model to an investigative framework of relevance for studying end-users in the contexts of tasks, problem at hand, knowledge, cognition, time, environment, uncertainty, judgments and evaluative responses.

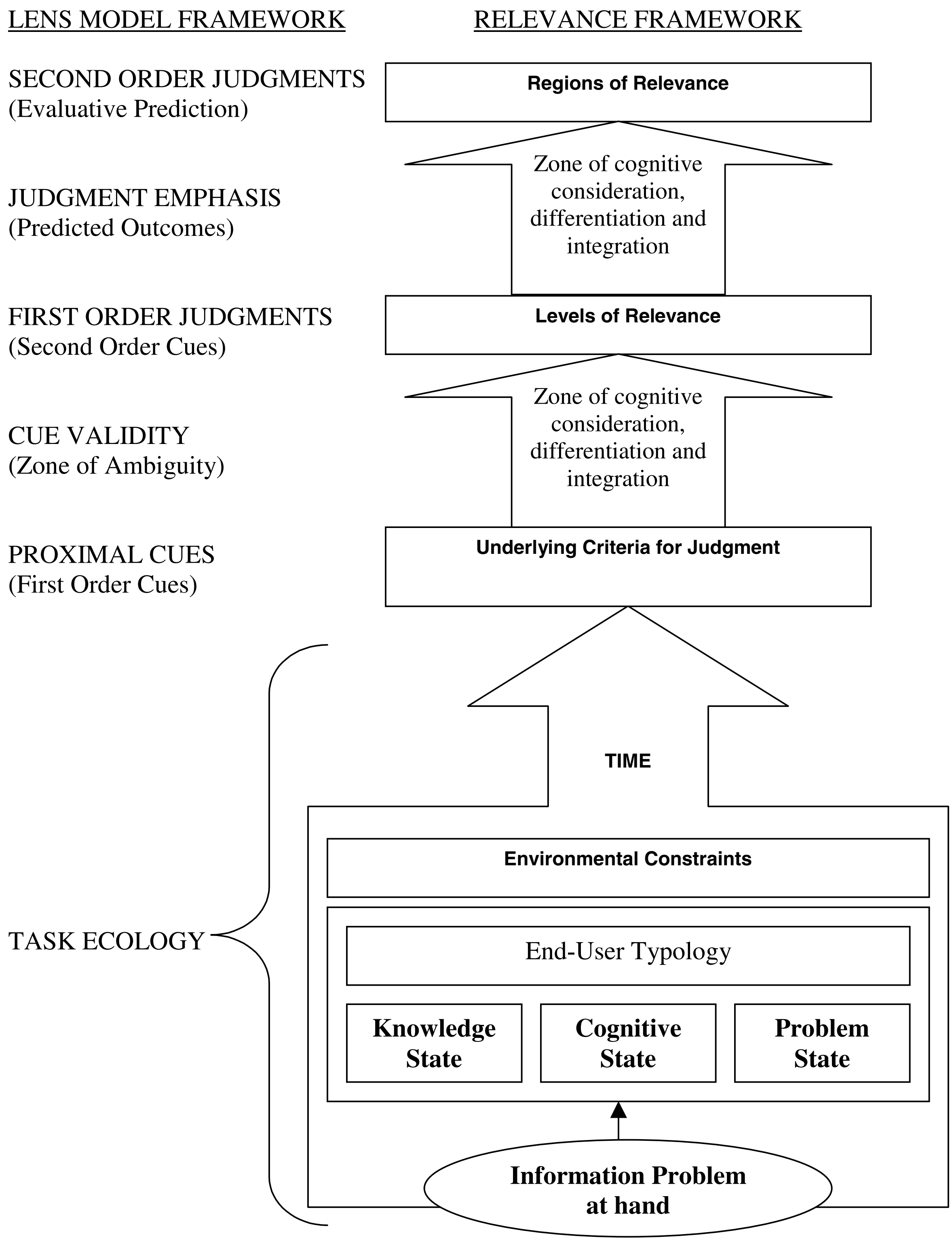

Figure 6.4 A Lens Model investigative framework for studying relevance

As posited by Saracevic (1999),

> The success or failure of any interactive system and technology is contingent on the extent to which user issues, the human factors, are addressed right from the beginning to the very end, right from theory, conceptualization, and design process on to development, evaluation, and to provision of services.

All of these human characteristics contribute theoretical perspectives surrounding relevance evaluations. The next best step is to take a next best step by moving beyond topicality as the major focus of IR system development. In the end, systems that can serve users by understanding users will emerge as the systems of choice, with the greatest effectiveness and the highest efficiency in the world of information.

APPENDIX A

RELEVANCE WORKSHEET

RELEVANCE WORKSHEET

	JUDGEMENTS (check one box only)					LEVELS OF RELEVANCE (check box(s) most important to your judgement)									
ITEM#	NR	PNR	PR	R		S	T	P	U	M	NS	NT	NP	NU	NM

APPENDIX B

END-USER INSTRUCTIONS

EVALUATION INSTRUCTIONS (for use with the attached form):

(1) Indicate your relevance judgement (check just one) in the boxes provided, identifying whether you consider the item to be not relevant [NR], partially not relevant [PNR], partially relevant [PR], or relevant [R];

(2) Identify the level of relevance by checking the box(s) that indicate the **most important reasons** for making the judgement as you did. These levels are defined as follows:

[S] SYSTEMATIC: The item retrieved was in a form/format that meets my information need;

[T] TOPICAL: The item retrieved was on the topic/subject I requested;

[P] PERTINENCE: I believe the item retrieved is/will be informative;

[U] UTILITY: The item retrieved is/will be useful in resolving my current/or a future information need;

[M] MOTIVATIONAL: The item retrieved will/may cause me to take other action(s) now that I have this information;

[NS] SYSTEMATIC: The item retrieved was NOT in a form/format that meets my information need;

[NT] TOPICAL: The item retrieved was NOT on the topic/subject I requested;

[NP] PERTINENCE: I believe the item retrieved is NOT/will NOT be informative;

[NU] UTILITY: The item retrieved is NOT/will NOT be useful in resolving my current/or a future information need;

[NM] MOTIVATIONAL: The item retrieved will/may NOT cause me to take other action(s) now that I have this information.

APPENDIX C

DATA CORPUS INCLUDING ALL POSSIBLE JUDGMENTS AND EVALUATIONS

Systematic	Topical	Pertinence	Utility	Motivation	NR	PNR	PR	R	Total
Positive	Positive	Positive	Positive	Positive			3	108	111
Positive	Positive	Positive	Positive	Negative		1	3	1	5
Positive	Positive	Positive	Positive	NC			3	59	62
Positive	Positive	Positive	Negative	Positive					0
Positive	Positive	Positive	Negative	Negative		2	4		6
Positive	Positive	Positive	Negative	NC			1		1
Positive	Positive	Positive	NC	Positive				5	5
Positive	Positive	Positive	NC	Negative			1		1
Positive	Positive	Positive	NC	NC			13	9	22
Positive	Positive	Negative	Positive	Positive					0
Positive	Positive	Negative	Positive	Negative					0
Positive	Positive	Negative	Positive	NC					0
Positive	Positive	Negative	Negative	Positive			2		2
Positive	Positive	Negative	Negative	Negative	3	18	1		22
Positive	Positive	Negative	Negative	NC					0
Positive	Positive	Negative	NC	Positive		1			1
Positive	Positive	Negative	NC	Negative					0
Positive	Positive	Negative	NC	NC					0
Positive	Positive	NC	Positive	Positive					0
Positive	Positive	NC	Positive	Negative					0
Positive	Positive	NC	Positive	NC			5	2	7
Positive	Positive	NC	Negative	Positive					0
Positive	Positive	NC	Negative	Negative		2	1		3
Positive	Positive	NC	Negative	NC	7	1			8
Positive	Positive	NC	NC	Positive		2	3	1	6
Positive	Positive	NC	NC	Negative					0
Positive	Positive	NC	NC	NC		1	2	2	5
Positive	Negative	Positive	Positive	Positive			1		1
Positive	Negative	Positive	Positive	Negative					0
Positive	Negative	Positive	Positive	NC					0
Positive	Negative	Positive	Negative	Positive					0
Positive	Negative	Positive	Negative	Negative					0
Positive	Negative	Positive	Negative	NC					0
Positive	Negative	Positive	NC	Positive			1		1
Positive	Negative	Positive	NC	Negative					0
Positive	Negative	Positive	NC	NC					0
Positive	Negative	Negative	Positive	Positive		1			1
Positive	Negative	Negative	Positive	Negative					0
Positive	Negative	Negative	Positive	NC	1				1
Positive	Negative	Negative	Negative	Positive					0
Positive	Negative	Negative	Negative	Negative					0
Positive	Negative	Negative	Negative	NC	1				1
Positive	Negative	Negative	NC	Positive					0
Positive	Negative	Negative	NC	Negative	1				1
Positive	Negative	Negative	NC	NC	1				1
Positive	Negative	NC	Positive	Positive					0
Positive	Negative	NC	Positive	Negative					0
Positive	Negative	NC	Positive	NC					0
Positive	Negative	NC	Negative	Positive	2	1			3

Positive	Negative	NC	Negative	Negative					0
Positive	Negative	NC	Negative	NC					0
Positive	Negative	NC	NC	Positive		1	2		3
Positive	Negative	NC	NC	Negative					0
Positive	Negative	NC	NC	NC		1			1
Positive	NC	Positive	Positive	Positive			3		3
Positive	NC	Positive	Positive	Negative					0
Positive	NC	Positive	Positive	NC			4	3	7
Positive	NC	Positive	Negative	Positive		1			1
Positive	NC	Positive	Negative	Negative		1			1
Positive	NC	Positive	Negative	NC			1		1
Positive	NC	Positive	NC	Positive		2	4	2	8
Positive	NC	Positive	NC	Negative					0
Positive	NC	Positive	NC	NC			4		4
Positive	NC	Negative	Positive	Positive					0
Positive	NC	Negative	Positive	Negative					0
Positive	NC	Negative	Positive	NC					0
Positive	NC	Negative	Negative	Positive					0
Positive	NC	Negative	Negative	Negative					0
Positive	NC	Negative	Negative	NC		1			1
Positive	NC	Negative	NC	Positive					0
Positive	NC	Negative	NC	Negative					0
Positive	NC	Negative	NC	NC		1			1
Positive	NC	NC	Positive	Positive			2	2	4
Positive	NC	NC	Positive	Negative					0
Positive	NC	NC	Positive	NC		9	15	1	25
Positive	NC	NC	Negative	Positive					0
Positive	NC	NC	Negative	Negative					0
Positive	NC	NC	Negative	NC		3			3
Positive	NC	NC	NC	Positive		1	4		5
Positive	NC	NC	NC	Negative					0
Positive	NC	NC	NC	NC			10	7	17
Negative	Positive	Positive	Positive	Positive					0
Negative	Positive	Positive	Positive	Negative					0
Negative	Positive	Positive	Positive	NC					0
Negative	Positive	Positive	Negative	Positive			1		1
Negative	Positive	Positive	Negative	Negative					0
Negative	Positive	Positive	Negative	NC					0
Negative	Positive	Positive	NC	Positive					0
Negative	Positive	Positive	NC	Negative					0
Negative	Positive	Positive	NC	NC					0
Negative	Positive	Negative	Positive	Positive					0
Negative	Positive	Negative	Positive	Negative					0
Negative	Positive	Negative	Positive	NC					0
Negative	Positive	Negative	Negative	Positive					0
Negative	Positive	Negative	Negative	Negative	15	9			24
Negative	Positive	Negative	Negative	NC					0
Negative	Positive	Negative	NC	Positive					0
Negative	Positive	Negative	NC	Negative					0
Negative	Positive	Negative	NC	NC					0

Negative	Positive	NC	Positive	Positive					0
Negative	Positive	NC	Positive	Negative					0
Negative	Positive	NC	Positive	NC					0
Negative	Positive	NC	Negative	Positive					0
Negative	Positive	NC	Negative	Negative					0
Negative	Positive	NC	Negative	NC					0
Negative	Positive	NC	NC	Positive					0
Negative	Positive	NC	NC	Negative					0
Negative	Positive	NC	NC	NC			1		1
Negative	Negative	Positive	Positive	Positive			1		1
Negative	Negative	Positive	Positive	Negative			1		1
Negative	Negative	Positive	Positive	NC					0
Negative	Negative	Positive	Negative	Positive					0
Negative	Negative	Positive	Negative	Negative	4				4
Negative	Negative	Positive	Negative	NC					0
Negative	Negative	Positive	NC	Positive					0
Negative	Negative	Positive	NC	Negative					0
Negative	Negative	Positive	NC	NC					0
Negative	Negative	Negative	Positive	Positive					0
Negative	Negative	Negative	Positive	Negative					0
Negative	Negative	Negative	Positive	NC					0
Negative	Negative	Negative	Negative	Positive					0
Negative	Negative	Negative	Negative	Negative	41				41
Negative	Negative	Negative	Negative	NC	2				2
Negative	Negative	Negative	NC	Positive					0
Negative	Negative	Negative	NC	Negative					0
Negative	Negative	Negative	NC	NC					0
Negative	Negative	NC	Positive	Positive					0
Negative	Negative	NC	Positive	Negative					0
Negative	Negative	NC	Positive	NC					0
Negative	Negative	NC	Negative	Positive					0
Negative	Negative	NC	Negative	Negative	2				2
Negative	Negative	NC	Negative	NC					0
Negative	Negative	NC	NC	Positive					0
Negative	Negative	NC	NC	Negative	1				1
Negative	Negative	NC	NC	NC	1				1
Negative	NC	Positive	Positive	Positive					0
Negative	NC	Positive	Positive	Negative					0
Negative	NC	Positive	Positive	NC					0
Negative	NC	Positive	Negative	Positive					0
Negative	NC	Positive	Negative	Negative					0
Negative	NC	Positive	Negative	NC					0
Negative	NC	Positive	NC	Positive					0
Negative	NC	Positive	NC	Negative					0
Negative	NC	Positive	NC	NC					0
Negative	NC	Negative	Positive	Positive					0
Negative	NC	Negative	Positive	Negative					0
Negative	NC	Negative	Positive	NC					0
Negative	NC	Negative	Negative	Positive					0
Negative	NC	Negative	Negative	Negative		1			1

Negative	NC	Negative	Negative	NC					0
Negative	NC	Negative	NC	Positive		1			1
Negative	NC	Negative	NC	Negative					0
Negative	NC	Negative	NC	NC		1			1
Negative	NC	NC	Positive	Positive					0
Negative	NC	NC	Positive	Negative					0
Negative	NC	NC	Positive	NC					0
Negative	NC	NC	Negative	Positive					0
Negative	NC	NC	Negative	Negative					0
Negative	NC	NC	Negative	NC	2				2
Negative	NC	NC	NC	Positive			2		2
Negative	NC	NC	NC	Negative					0
Negative	NC	NC	NC	NC	11	6	4		21
NC	Positive	Positive	Positive	Positive				5	5
NC	Positive	Positive	Positive	Negative					0
NC	Positive	Positive	Positive	NC			3	51	54
NC	Positive	Positive	Negative	Positive			1		1
NC	Positive	Positive	Negative	Negative					0
NC	Positive	Positive	Negative	NC		1	32		33
NC	Positive	Positive	NC	Positive				2	2
NC	Positive	Positive	NC	Negative					0
NC	Positive	Positive	NC	NC			14	42	56
NC	Positive	Negative	Positive	Positive					0
NC	Positive	Negative	Positive	Negative					0
NC	Positive	Negative	Positive	NC			19		19
NC	Positive	Negative	Negative	Positive					0
NC	Positive	Negative	Negative	Negative	1	1			2
NC	Positive	Negative	Negative	NC		4			4
NC	Positive	Negative	NC	Positive					0
NC	Positive	Negative	NC	Negative					0
NC	Positive	Negative	NC	NC		12	3	1	16
NC	Positive	NC	Positive	Positive				2	2
NC	Positive	NC	Positive	Negative					0
NC	Positive	NC	Positive	NC			3	5	8
NC	Positive	NC	Negative	Positive					0
NC	Positive	NC	Negative	Negative					0
NC	Positive	NC	Negative	NC	52	2	8		62
NC	Positive	NC	NC	Positive				2	2
NC	Positive	NC	NC	Negative					0
NC	Positive	NC	NC	NC	1	1	27	16	45
NC	Negative	Positive	Positive	Positive					0
NC	Negative	Positive	Positive	Negative					0
NC	Negative	Positive	Positive	NC					0
NC	Negative	Positive	Negative	Positive		1	1		2
NC	Negative	Positive	Negative	Negative	2				2
NC	Negative	Positive	Negative	NC		1			1
NC	Negative	Positive	NC	Positive			1		1
NC	Negative	Positive	NC	Negative					0
NC	Negative	Positive	NC	NC		9			9
NC	Negative	Negative	Positive	Positive					0

NC	Negative	Negative	Positive	Negative					0
NC	Negative	Negative	Positive	NC					0
NC	Negative	Negative	Negative	Positive					0
NC	Negative	Negative	Negative	Negative	49	1	1		51
NC	Negative	Negative	Negative	NC	7				7
NC	Negative	Negative	NC	Positive					0
NC	Negative	Negative	NC	Negative	2				2
NC	Negative	Negative	NC	NC	15	1			16
NC	Negative	NC	Positive	Positive					0
NC	Negative	NC	Positive	Negative					0
NC	Negative	NC	Positive	NC			3		3
NC	Negative	NC	Negative	Positive					0
NC	Negative	NC	Negative	Negative	5				5
NC	Negative	NC	Negative	NC	11	3			14
NC	Negative	NC	NC	Positive		3			3
NC	Negative	NC	NC	Negative					0
NC	Negative	NC	NC	NC	109	20	3		132
NC	NC	Positive	Positive	Positive				1	1
NC	NC	Positive	Positive	Negative					0
NC	NC	Positive	Positive	NC		2	4	2	8
NC	NC	Positive	Negative	Positive					0
NC	NC	Positive	Negative	Negative		1			1
NC	NC	Positive	Negative	NC		3	1		4
NC	NC	Positive	NC	Positive			2	2	4
NC	NC	Positive	NC	Negative					0
NC	NC	Positive	NC	NC			29	21	50
NC	NC	Negative	Positive	Positive					0
NC	NC	Negative	Positive	Negative					0
NC	NC	Negative	Positive	NC		1			1
NC	NC	Negative	Negative	Positive					0
NC	NC	Negative	Negative	Negative	29				29
NC	NC	Negative	Negative	NC	25	3			28
NC	NC	Negative	NC	Positive					0
NC	NC	Negative	NC	Negative	18				18
NC	NC	Negative	NC	NC	59	60	5		124
NC	NC	NC	Positive	Positive			2		2
NC	NC	NC	Positive	Negative					0
NC	NC	NC	Positive	NC			21	22	43
NC	NC	NC	Negative	Positive					0
NC	NC	NC	Negative	Negative	6	1			7
NC	NC	NC	Negative	NC	54	13			67
NC	NC	NC	NC	Positive		2	3	6	11
NC	NC	NC	NC	Negative	1	4	1		6
NC	NC	NC	NC	NC					0
				Total	**541**	**219**	**290**	**382**	**1432**

APPENDIX D

DATA CORPUS OF ACTUAL END-USER JUDGMENTS AND EVALUATIONS

Systematic	Topical	Pertinence	Utility	Motivation	NR	PNR	PR	R	Total
NC	Negative	NC	NC	NC	109	20	3		132
NC	NC	Negative	NC	NC	59	60	5		124
Positive	Positive	Positive	Positive	Positive			3	108	111
NC	NC	NC	Negative	NC	54	13			67
Positive	Positive	Positive	Positive	NC			3	59	62
NC	Positive	NC	Negative	NC	52	2	8		62
NC	Positive	Positive	NC	NC			14	42	56
NC	Positive	Positive	Positive	NC			3	51	54
NC	Negative	Negative	Negative	Negative	49	1	1		51
NC	NC	Positive	NC	NC			29	21	50
NC	Positive	NC	NC	NC	1	1	27	16	45
NC	NC	NC	Positive	NC			21	22	43
Negative	Negative	Negative	Negative	Negative	41				41
NC	Positive	Positive	Negative	NC		1	32		33
NC	NC	Negative	Negative	Negative	29				29
NC	NC	Negative	Negative	NC	25	3			28
Positive	NC	NC	Positive	NC		9	15	1	25
Negative	Positive	Negative	Negative	Negative	15	9			24
Positive	Positive	Positive	NC	NC			13	9	22
Positive	Positive	Negative	Negative	Negative	3	18	1		22
Negative	NC	NC	NC	NC	11	6	4		21
NC	Positive	Negative	Positive	NC			19		19
NC	NC	Negative	NC	Negative	18				18
Positive	NC	NC	NC	NC			10	7	17
NC	Positive	Negative	NC	NC		12	3	1	16
NC	Negative	Negative	NC	NC	15	1			16
NC	Negative	NC	Negative	NC	11	3			14
NC	NC	NC	NC	Positive		2	3	6	11
NC	Negative	Positive	NC	NC		9			9
Positive	Positive	NC	Negative	NC	7	1			8
Positive	NC	Positive	NC	Positive		2	4	2	8
NC	Positive	NC	Positive	NC			3	5	8
NC	NC	Positive	Positive	NC		2	4	2	8
Positive	Positive	NC	Positive	NC			5	2	7
Positive	NC	Positive	Positive	NC			4	3	7
NC	Negative	Negative	Negative	NC	7				7
NC	NC	NC	Negative	Negative	6	1			7
Positive	Positive	Positive	Negative	Negative		2	4		6
Positive	Positive	NC	NC	Positive		2	3	1	6
NC	NC	NC	NC	Negative	1	4	1		6
Positive	Positive	Positive	Positive	Negative		1	3	1	5
Positive	Positive	Positive	NC	Positive				5	5
Positive	Positive	NC	NC	NC		1	2	2	5
Positive	NC	NC	NC	Positive		1	4		5
NC	Positive	Positive	Positive	Positive				5	5
NC	Negative	NC	Negative	Negative	5				5
Positive	NC	Positive	NC	NC			4		4
Positive	NC	NC	Positive	Positive			2	2	4
Negative	Negative	Positive	Negative	Negative	4				4

NC	Positive	Negative	Negative	NC		4			4
NC	NC	Positive	Negative	NC		3	1		4
NC	NC	Positive	NC	Positive			2	2	4
Positive	Positive	NC	Negative	Negative		2	1		3
Positive	Negative	NC	Negative	Positive	2	1			3
Positive	Negative	NC	NC	Positive		1	2		3
Positive	NC	Positive	Positive	Positive			3		3
Positive	NC	NC	Negative	NC		3			3
NC	Negative	NC	Positive	NC			3		3
NC	Negative	NC	NC	Positive		3			3
Positive	Positive	Negative	Negative	Positive			2		2
Negative	Negative	Negative	Negative	NC	2				2
Negative	Negative	NC	Negative	Negative	2				2
Negative	NC	NC	Negative	NC	2				2
Negative	NC	NC	NC	Positive			2		2
NC	Positive	Positive	NC	Positive				2	2
NC	Positive	Negative	Negative	Negative	1	1			2
NC	Positive	NC	Positive	Positive				2	2
NC	Positive	NC	NC	Positive				2	2
NC	Negative	Positive	Negative	Positive		1	1		2
NC	Negative	Positive	Negative	Negative	2				2
NC	Negative	Negative	NC	Negative	2				2
NC	NC	NC	Positive	Positive			2		2
Positive	Positive	Positive	Negative	NC			1		1
Positive	Positive	Positive	NC	Negative			1		1
Positive	Positive	Negative	NC	Positive		1			1
Positive	Negative	Positive	Positive	Positive			1		1
Positive	Negative	Positive	NC	Positive			1		1
Positive	Negative	Negative	Positive	Positive		1			1
Positive	Negative	Negative	Positive	NC	1				1
Positive	Negative	Negative	Negative	NC	1				1
Positive	Negative	Negative	NC	Negative	1				1
Positive	Negative	Negative	NC	NC	1				1
Positive	Negative	NC	NC	NC		1			1
Positive	NC	Positive	Negative	Positive		1			1
Positive	NC	Positive	Negative	Negative		1			1
Positive	NC	Positive	Negative	NC			1		1
Positive	NC	Negative	Negative	NC		1			1
Positive	NC	Negative	NC	NC		1			1
Negative	Positive	Positive	Negative	Positive			1		1
Negative	Positive	NC	NC	NC			1		1
Negative	Negative	Positive	Positive	Positive			1		1
Negative	Negative	Positive	Positive	Negative			1		1
Negative	Negative	NC	NC	Negative	1				1
Negative	Negative	NC	NC	NC	1				1
Negative	NC	Negative	Negative	Negative		1			1
Negative	NC	Negative	NC	Positive		1			1
Negative	NC	Negative	NC	NC		1			1
NC	Positive	Positive	Negative	Positive			1		1
NC	Negative	Positive	Negative	NC		1			1

NC	Negative	Positive	NC	Positive			1		1
NC	NC	Positive	Positive	Positive				1	1
NC	NC	Positive	Negative	Negative		1			1
NC	NC	Negative	Positive	NC		1			1
				Total	**541**	**219**	**290**	**382**	**1432**

REFERENCES

Abu-Mostafa, Y. S. (1986). Complexity in Information Theory. New York, NY: Springer Verlag.

Adelman, L. (1981). The influence of formal, substantive, and contextual task properties on the relative effectiveness of different forms of feedback in multiple-cue probability learning tasks. Organizational Behavior and Human Performance, 27, 423-442.

Agnes, M. (1996). Webster's New World Dictionary and Thesaurus. New York, NY:Simon & Schuster, Inc.

Ajzen, I. (1989). Attitude structure and behavior. A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude Structure and Function (pp. 241-274). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Allen, B. L. (1991). Cognitive research in information science: Implications for design. Annual Review of Information Science and Technology (ARIST), 26, 3-37.

Arkes, H. R., & Hammond, K. R. (1986). General introduction. H. R. Arkes, & K. R. Hammond (Eds.), Judgment and Decision Making: An Interdisciplinary Reader (pp. 1-11). Cambridge, MA:Cambridge University Press.

Aschenbrenner, M., Bockenholt, U., Albert, D., & Schmalhofer, F. (1986). The selection of dimensions when choosing between multiattribute alternatives. R. W. Scholz (Ed.), Current Issues in West German decision research (pp. 63-78). Frankfurt, Germany.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. Acta

Psychology, 44, 211-233.

Barry, C. L. (1993). The identification of user relevance criteria and document characteristics: Beyond the topical approach to information retrieval. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.

Barry, C. L. (1994). User-defined relevance criteria: An exploratory study. Journal of the American Society for Information Science, 45(3), 149-159.

Barry, C. L., & Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. Information Processing & Management, 34(2 3), 219-236.

Bateman, J. A. (1998). Modeling changes in end-user relevance criteria: An Information seeking study. Unpublished doctoral dissertation, University of North Texas, Denton, Texas.

Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening options. Psychological Science, 4, 215-220.

Bearman, D. (1997). Research issues in metadata. Working Meeting on Electronic Records Research. May 28-30, Pittsburgh, PA. [Online] Available: http//:www.sis.pitt.edu/~cerar/s4-db.htm.

Belkin, N. J. (1978). Progress in documentation: Information concepts for information science. Journal of Documentation, 34(1), 55-85.

Berlyne, D. E. (1960). Conflict, arousal and curiosity. New York, NY: McGraw-Hill.

Blair, D. C. (1990). Language and Representation in Information Retrieval. New York, NY: Elsevier Science Publishing Company, Inc.

Blair, D. C. (1996). STAIRS redux: Thoughts on the STAIRS evaluation, ten years after. Journal of the American Society of Information Science, 47(1), 4-22.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. Science, 236, 157-161.

Brooks, T. A. (1995). People, words, and perceptions: A phenomenological investigation of textuality. Journal of the American Society for Information Science, 46(2), 103-115.

Brooks, T. A. (1997). The relevance aura of bibliographic records. Information Processing & Management, 33(1), 69-80.

Bruce, H. W. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. Journal of the American Society for Information Science, 45(3), 142-148.

Brunswik, E. (1952). The Conceptual Framework of Psychology. Chicago, IL: University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. Psychological Review, 62, 193-217.

Brunswik, E. (1956). Perception and the Representative Design of Psychological Experiments. Berkeley, CA: University of California Press.

Carroll, J. S., & Johnson, E. J. (1990). Decision Research: A Field Guide. Newbury Park,CA: Sage Publications, Inc.

Cook, R. L. (1976). A study of interactive judgment analysis and the representation of weights in judgment policies. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

Cooksey, R. W. (1996). Judgment Analysis: Theory, methods, and applications. San Diego, CA: Academic Press, Inc.

Cool, C., Belkin, N. J., Frieder, O., & Kantor, P. (1993). Characteristics of texts affecting relevance judgments. Proceedings of the 14th National Online Meeting Medford, NJ: Learned Information, Inc.

Cooper, W. S. (1971). A definition of relevance for information retrieval. Information Storage & Retrieval, 7, 19-37.

Cooper, W. S. (1978). A perspective on the measurement of retrieval effectiveness. Drexel Library Journal, 14(2), 25-39.

Cosijn, E., & Ingwersen, P. (In Press). Dimensions of relevance. 1-24.

Crestini, F., Lalmas, M., & Van Rijsbergen, C. J. (1998). Information Retrieval: Uncertainty and Logics. Boston, MA: Kluwer Academic Publishers.

Daniels, P. J. (1986). Cognitive models in information retrieval - an evaluative review. Journal of Documentation, 42(4), 272-304.

DeMey, M. (1980). The relevance of the cognitive paradigm for information science. O. Harbo (Ed.), Theory and Application of Information Research (pp. 48-61). London, England: Mansell.

Dervin, B. (1983). An overview of sense-making research: Concepts, methods, and results to date. International Communication Association Annual Meeting .

Dervin, B., & Nilan, M. (1986). Information needs and uses. M. E. Williams

(Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 21pp. 3-33). Knowledge Industry Publications, Inc.

Dretske, F. I. (1981). Knowledge & the Flow of Information. Cambridge, MA: The MIT Press.

Eagly, A. H., & Chaiken, S. (1993). The Psychology of Attitudes. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Ebenbach, D. H., & Moore, C. F. (2000). Incomplete information, inferences, and individual differences: The case of environmental judgments. Organizational Behavior and Human Decision Processes, 81(1), 1-27.

Einhorn, H. J. (1978). Learning from experience and suboptimal rules in decision making. Cognitive Processes in Choice and Decision Behavior: Proceedings of a conference held under Contract N00014-78-C-0170 issued by the Office of Naval Research (pp. 1-17).

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 32, 53-88.

Eisenberg, M. B. (1988). Measuring relevance judgments. Information Processing & Management, 24(4), 373-389.

Ellis, D. (1989). A behavioral approach to information retrieval system design. Journal of Documentation, 45(3), 171-212.

Ellis, D. (1992). The physical and cognitive paradigms in information retrieval research. Journal of Documentation, 48(1), 45-64.

Ellis, D. (1996). The dilemma of measurement in information retrieval research.

Journal of the American Society for Information Science, 47(1), 23-36.

Fairthorne, R. A. (1958). Summary of Area 6 Discussion. Proceedings of the International Conference in Scientific Information, Washington, DC, 1406.

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.

Foskett, D. J. (1972). A note on the concept of "relevance". Information Storage & Retrieval, 8, 77-78.

Frants, V. I., & Brush, C. B. (1988). The need for information and some aspects of information retrieval systems construction. Journal of the American Society for Information Science, 39(2), 86-91.

Froehlich, T. J. (1991). Towards a better conceptual framework for understanding relevance for information science research. ASIS '91: Proceedings of the 54th Annual Meeting of the American Society for Information Science (pp. 118-125). Medford, NJ: Learned Information.

Froehlich, T. J. (1994). Relevance reconsidered -Towards an agenda for the 21st century: Introduction to special topic issue on relevance research. Journal of the American Society for Information Science, 45(3), 124-134.

Gagne, R. M. (1985). The Conditions of Learning (4th Edition). New York, NY: Holt, Rinehart and Winston.

Gluck, M. (1996). Exploring the relationship between user satisfaction and relevance in information systems. Information Processing & Management, 32(1), 89-104.

Goffman. (1965). On relevance as a measure. Information Storage & Retrieval, 2, 201-203.

Greisdorf, H., & Spink, A. (1999). Regions of relevance: Approaches to measurement for enhanced precision. Proceedings of the 21st British Computer Society Information Retrieval Sub Group Annual Colloquium on IR Research (pp. 1-33).

Hammond, K. R. (1955). Probabilistic functionalism and the clinical method. Psychological Review, 62, 255-262.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. (1975). Social judgment theory. M. Kaplan, & S. Schwartz (Eds.), Human Judgment and Decision Processes (pp. 271-312). New York, NY: Academic Press.

Harter, S. P. (1992). Psychological relevance and information science. Journal of the American Society for Information Science, 43(9), 602-615.

Harter, S. P. (1996). Variations in relevance assessments and the measurement of retrieval effectiveness. Journal of the American Society for Information Science, 47(1), 37-49.

Hayes, R. M. (1993). Measurement of information. Information Processing & Management, 29(1), 1-11.

Hillman, D. J. (1964). The notion of relevance. American Documentation, January, 26-34.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1994). Applied Statistics for the Behavioral Sciences. Boston, MA: Houghton Mifflin Company.

Howard, D. L. (1994). Pertinence as reflected in personal constructs. Journal of the American Society for Information Science, 45(3), 172-185.

Howell, W. C., & Burnett, S. A. (1978). Uncertainty measurement: A cognitive taxonomy. Organizational Behavior and Human Performance, 22, 45-68.

Ingwersen, P. (1982). Search procedures in the library - analysed from the cognitive viewpoint. Journal of Documentation, 38(3), 165-191.

Ingwersen, P. (1987). Towards a new research paradigm in information retrieval. I. Wormell (Ed.), Knowledge Engineering: Expert Systems and Information Retrieval (pp. 150-168). London, England: Taylor Graham.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. Journal of Documentation, 52(1), 3-50.

Ingwersen, P., & Borlund, P. Information transfer viewed as interactive cognitive processes. Second International Conference on Conceptions of Library and Information Science: Integration in Perspective (pp. 219-232).

Janes, J. W. (1994). Other poeple's judgments: A comparison of users' and others' judgments of document relevance, topicality and utility. Journal of the American Society for Information Science, 45(3), 160-171.

Janis, I. L. (1977). Decision Making. New York, NY: The Free Press.

Katter, R. V. (1968). The influence of scale form on relevance judgments. Information Storage & Retrieval, 4, 1-11.

Kelly, G. A. (1955). The Psychology of Personal Constructs. New York, NY: W. W. Norton & Company, Inc.

Kelly, G. A. (1969). A mathematical approach to psychology. (pp. 94-113). New

York: Wiley.

Kent, A., Berry, M., Leuhrs, F. U., & Perry, J. W. (1955). Machine literature searching VIII: Operational criteria for designing information retrieval systems. American Documentation, 6 (2), 93-101.

Kerlinger, F. N. (1967). Foundations of Behavioral Research. New York, NY: Holt, Rinehart and Winston, Inc.

Kreitler, H., & Kreitler, S. (1976). Cognitive Orientation and Behavior. New York, NY: Springer Publishing Company.

Lalmas, M. (1998). The flow of informatin in information retrieval: Towards a general framework for the modelling of information retrieval. F. Crestani, & M. &. v. R. C. J. Lalmas Information Retrieval: Uncertainty and Logics . Boston, MA: Kluwer Academic Publishers.

Linville, P. W. (1982). The complexity-extremity effect and age-based stereotyping. Journal of Personality and Social Psychology, 42(2), 193-211.

MacMullin, S. E., & Taylor, R. S. (1984). Problem Dimensions and information traits. Information Society, 3(1), 91-111.

Marchionini, G. (1995). Information seeking in electronic environments. Cambridge, UK: Cambridge Universitiy Press.

Maron, M. E. (1978). Theory and foundations of information retrieval: Some introductory remarks. Drexel Library Quarterly, 14(2), 1-9.

McGuire, W. J. (1989). The structure of individual attitudes and attitude systems. A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude Structure and

Function (pp. 37-69). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Medin, D. (1976). Theories of discrimination learning and learning set. (Vol. Handbook of Learning and Cognitive Processespp. Chapter 4, 131-169). Hillside, NJ: Lawrence Earlbaum Associates, Publishers.

Meghini, C., Sebastiani, F., & Straccia, U. (1998). MIRLOG: A logic for multimedia information retrieval. F. Crestini, M. Lalmas, & C. J. van Rijsbergen (Eds.), Information Retrieval: Uncertainty and Logics (pp. 151-185). Boston, MA: Kluwer Academic Publishers.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Miller, P. M. (1971). Do labels mislead? A multiple cue study, within the framework of Brunswik's probabilistic functionalism. Organizational Behavior and Human Performance, 6, 480-500.

Mizzaro, S. (1996). A cognitive analysis of information retrieval. Second International Conference on Conceptions of Library and Information Science: Integration in Perspective (CoLIS 2) (pp. 233-250).

Mizzaro, S. (1998). How many relevances in information retrieval? Interacting With Computers, 10, 303-320.

Montgomery, H., & Svenson, O. (1976). On decision rules and information processing strategies for choices among multiattribute alternatives. Scandinavian Journal of Psychology, 17, 283-291.

Montgomery, H., & Willen, H. (1999). Decision making and action: The search for a good structure. Peter Juslin, & Henry Montgomery (Eds.), Judgment and Decision Making: Neo-Brunswickian and Process-Tracing Approaches (pp. 147-173). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, MA: Harvard University Press.

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

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal report on mental processes. Psychological Review, 84, 231-259.

Ostrom, T. M., & Upshaw, H. S. (1968). Psychological perspective and attitude change. A. Greenwald, T. Brock, & T. Ostrom (Eds.), Psychological Foundations of Attitudes (pp. 217-242). New York, NY: Academic Press.

Ostrom, T. (1989). Interdependence of attitude theory and measurement. A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude Structure and Function (pp. 11-36). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Overton, Willis F. (1990). Reasoning, necessity, and logic: Developmental Perspectives. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Park, H. (1997). Relevance of science information: Origins and dimensions of relevance and their implications to information retrieval. Information Processing & Management, 33(3), 339-352.

Park, T. K. (1993). The nature of relevance in information retrieval: An empirical study. Library Quaterly, 63(3), 318-351.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance, 16, 366-387.

Pratkanis, A. R. (1989). The cognitive representation of attitudes. A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude Structure and Function (pp. 71-98). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Pribram, K. H. (1971). Languages of the Brain: Experimental paradoxes and principles in neuropsychology. Englewood Cliffs, NJ: Prentice-Hall.

Rees, A. M. (1966). The relevance of relevance to the testing and evaluation of document retrieval systems. Aslib Proceedings, 18(11), 316-324.

Robertson, S. E., Maron, M. E., & Cooper, W. S. (1982). Probability of relevance: A unification of two competing models for document retrieval. Information Technology: Research and Development, 1, 1-21.

Robins, D. B. (1998). Shifts of focus among dimensions of user information problems as represented during interactive information retrieval. Unpublished doctoral dissertation, University of North Texas, Denton, TX.

Rosen, L. D., & Rosenkoetter, P. (1976). An eye fixation analysis of choice and judgment with multiattribute stimuli. Memory and Cognition, 4, 747-752.

Saracevic, T. (1988). A study of information seeking and retrieving. 1. Background and methodology. Journal of the American Society for Information Science, 39(3), 161-176.

Saracevic, T. (1995). Evaluation of evaluation in information retrieval.

Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval .

Saracevic, T. (1996). Relevance reconsidered '96. Second International Conference on Conceptions of Library and Information Science: Integration in Perspective (pp. 201-216).

Saracevic, T. (1999). Information science. Journal of the American Society for Information Science, 50(12), 1051-1063.

Schamber, L. (1991). Users' criteria for evaluation in multimedia information seeking and use situations. ASIS '91: Proceedings of the American Society for Information Science 54th Annual Meeting (pp. 126-133). Medford, NJ: Learned Information, Inc.

Schamber, L. (1994). Relevance and information behavior. Martha E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 29pp. 3-48). Medford, NJ: Learned Information, Inc.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. Information Processing & Management, 26(6), 755-776.

Schutz. (1970). Reflections on the Problem of Relevance. Westport, CT: Greenwood Press, Publishers.

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. Sociometry, 25, 405-414.

Scott, W. A. (1969). Structure of natural cognitions. Journal of Personality and

Social Psychology, 12 , 261-278.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. Symposium on the mechanisation of thought processes London, England: HM Stationery Office.

Settle, T. (1976). Induction and probability unfused. (Vol. The Philosophy of Karl Popper). La Salle, IL: Open Court.

Shaw, M. L. G., & Gaines, B. R. (1992) Kelly's "Geometry of Psychological Space" and its Significance for Cognitive Modeling [Online]. Available: http://ksi.cpsc.ucalgary.ca/articles/ NewPsych92/#References.

Sherif, C. W., Sherif, M., & Nebergall, R. E. (1965). Attitude and Attitude Change: The social judgment-involvement approach. Philadelphia, PA: Saunders.

Simon, H. A. (1955). A behavioral model of rational choice. Quarterly Journal of Economics, 59, 99-118.

Simon, H. A. (1972). Theories of bounded rationality. C. B. Radner, & R. Radner (Eds.), Decision and Organization . Amsterdam: North-Holland.

Sjoberg, L. (1999). Attitudes, values and opinions: Models and dynamics. Peter Juslin, & Henry Montgomery (Eds.), Judgment and Decision Making: Neo-Brunswickian and Process-Tracing Approaches (pp. 219-241). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Slovic, P. (1972). From Shakespeare to Simon: Speculation -- and some evidence about man's ability to process information. Oregon Research Institute Research Monograph, 12(12).

Smith, M. B., Bruner, J. S., & White, R. W. (1957). Opinions and Personality.

New York, NY: Wiley.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and Cognition. Cambridge, MA: Harvard University Press.

Spink, A., Bateman, J., & Greisdorf, H. (1999). Successive searching behavior during information seeking: An exploratory study. Journal of Information Science, 25(6), 439-449.

Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to nonrelevant: Examining different regions of relevance. Information Processing & Management, 34(5), 599-622.

Spink, A., & Greisdorf, H. (1999). How and why end-users make relevance judgments. Proceedings of the 20th National Online Meeting (pp. 239-250). New York, NY: Information Today, Inc.

Spink, A., & Greisdorf, H. (1997). Partial relevance judgments and changes in users information problems during online searching. Proceedings of the 18th National Online Meeting (pp. 323-334). New York, NY: Information Today, Inc.

Stewart, T. R. (1988). Judgment analysis: Procedures. B. Brehmer, & C. R. B. Joyce (Eds.), Human Judgment: The SJT view (pp. 41-74). Amsterdam: North-Holland Elsevier.

Su, L. T. (1994). The relevance of recall and precision in user evaluation. Journal of the American Society for Information Science, April, 207-217.

Sudman, S., & Bradburn, N. (1982). Asking questions: A practical guide to questionaire design. San Francisco, CA: Jossey-Bass.

Sutherland, N. S. (1969). Outlines of a theory of visual pattern recognition in animals and man. R. M. Gilbert, & N. S. Sutherland (Eds.), Animal discrimination learning . New York, NY: Academic Press.

Svenson, O. (1979). Process descriptions of decision making. Organizational Behavior and Human Performance, 23, 86-112.

Svenson, O. (1999). Differentiation and consolidation theory: Decision making processes before and after choice. Peter Juslin, & Henry Montgomery (Eds.), Judgment and Decision Making: Neo-Brunswickian and Process-Tracing Approaches (pp. 175-197). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Tague, J., & Schultz, R. (1989). Evaluation of the user interface in an information retrieval system: A model. Information Processing & Management, 25(4), 377-389.

Tang, R., Shaw, W. M. Jr., & Vevea, J. L. (1999). Towards the identification of the optimal number of relevance categories. Journal of the American Society for Information Science, 50 (3), 254-264.

Tang, R., & Solomon, P. (1998). Toward an understanding of the dynamics of relevance judgment: An analysis of one person's search behavior. Information Processing & Management, 34(2 3), 237-256.

Taube, M. (1965). A note on the pseudo-mathematics of relevance. American Documentation, 16(2), 69-72.

Thurstone, L. L. (1931). The measurement of attitudes. Journal of Abnormal and Social Psychology, 26, 249-269.

Turner, M. K. (1992). A study regarding information-seeking behavior, valuation of information including perceptions of information attributes, and associated correlates pertaining to information usage. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological Review, 79(4), 281-299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A. &. Kahneman. D. (1985). The framing of decisions and psychology of choice. George Wright (Ed.), Behavioral Decision Making. New York, NY: Plenum Press.

van der Pligt, J., & Eiser, R. J. (1984). Dimensional salience, judgment, and attitudes. R. J. Eiser (Ed.), Attitudinal Judgment (pp. 161-177). New York, NY: Springer-Verlag.

van Rijsbergen, C. J. (1975). Information Retrieval. London, England: Butterworths.

van Rijsbergen, C. J. (1979). Information Retrieval, 2nd Ed. London, England: Butterworths

van Rijsbergen, C. J. (1996). Information, logic, and uncertainty in information science. Second International Conference on Conceptions of Library and Information Science: Integration in Perspective (pp. 1-10).

van Rijsbergen, C. J. (1999). A non-classical logic for information retrieval. F. Crestini, M. Lalmas, & C. J. van Rijsbergen (Eds.), Information Retrieval: Uncertainty and Logics (pp. 3-13). Boston, MA: Kluwer Academic Publishers.

Vygotsky, L. S. (1978). Mind In Society. Cambridge, MA: Harvard University Press.

Wallis, P., & Thom, J. A. (1996). Relevance judgments for assessing recall. Information Processing & Management, 32(3), 273-286.

Wallsten, T. S. (1978). Processes and models to describe choice and inference behavior. Cognitive Processes in Choice and Decision Behavior: Proceedings of a conference held under Contract N00014-78-C-0170 issued by the Office of Naval Research (pp. 215-234).

Wallsten, T. S. (1980). Processes and models to describe choice and inference behavior. T. S. Wallsten (Ed.), Cognitive Processes in Choice and Decision Behavior (pp. 215-237). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wang, P. (1994). A cognitive model of document selection of real users of IR systems. Unpublished doctoral dissertation, University of Maryland, College Park, MD.

Wang, P. (1997). The design of document retrieval systems for academic users: Implications of studies on users' relevance criteria. Proceedings of the 60th Annual Meeting of the American Society for Information Science (pp. 162-173).

Wang, P., & Soergel, D. (1993). Beyond topical relevance: Document selection behavior of real users of IR systems. Proceedings of the 56th ASIS Annual Meeting (pp. 87-92). Medford, NJ: Learned Information, Inc.

Watzlawick, P., Bavelas, J. B., & Jackson, D. D. Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes. New

York, NY: W. W. Norton & Company.

Weibel, S. (1995). Metadata: The foundations of resource description. D-Lib Magazine, July. [Online] Available: http//:www.dlib.org/dlib/July95/07weibel.html.

Wilson, P. (1968). Two Kinds of Power: An Essay on Bibliographic Control. Berkeley, CA: University of California Press.

Wilson, P. (1977). Public Knowledge, Private Ignorance. Westport, CT: Greenwood Press, Inc.

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

Wilson, T. (1999). Models in information behavior research. Journal of Documentation, 55(3), 249-270.

Winograd, T., & Flores, F. (1987). Understanding Computers and Cognition. Reading, MA: Addison-Wesley Publishing Company.

Zajonc, R. B. (1960). The process of cognitive tuning in communication. Journal of Abnormal and Social Psychology, 61, 159-167.

evaluations of retrieved items, they appear to make those decisions by using consistent heuristic approaches.