Predicting the Utility of Feedback Judgements Using Cognitive Load Theory

By

Jonathan Back

A Doctoral Thesis submitted in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy of Loughborough University

September 2003

Supervisor: Professor Charles Oppenheim

Department of Information Science
Loughborough University

ABSTRACT

Results from laboratory testing suggest that user-based relevance feedback can significantly improve retrieval performance. However outside the laboratory, feedback systems are rarely utilised when implemented. This thesis explores why users are often reluctant to provide feedback. Modelling interaction involves reconciling the need for prediction with the seemingly individual-specific effect of information. Information behaviour is guided by heuristics and not by logical analysis or deduction. Heuristics impose assumptions that are used to address a problem in a way that is compatible with an individual's knowledge schemata.

This thesis argues that feedback heuristics are influenced by the cognitive load imposed on an individual. Cognitive load is defined as the intrinsic (cognitive) and extraneous (situational) difficulty associated with providing feedback during the search process. The application of Cognitive Load Theory has allowed factors identified by qualitative research as having an affect on behaviour to be explored. These factors have been used to develop a quantitative model capable of predicting the utility of feedback judgements.

Knowledge obtained from previous search interactions is responsible for intrinsic cognitive load, and consequently, could be partially responsible for the way individuals assimilate information. Individuals who were able to manage uncertainty in response to situational demands experienced less intrinsic cognitive load. Task complexity is responsible for extraneous cognitive load, and consequently, could be partially responsible for the way individuals interact with information objects. Selecting suitable heuristics that can in some way anticipate task needs reduces extraneous cognitive load. This research provides information science with a model that enables an improved understanding of how cognitive limitations influence interaction.

ACKNOWLEDGEMENTS

Firstly, I wish to thank Derek Stephens and Ron Summers. Derek provided me with the opportunity to study at Loughborough as an undergraduate. Ron's efforts ensured that I stayed at Loughborough after obtaining my B.Sc. Over the years I have developed a tremendous amount of respect for all of the staff within the Department of Information Science. I wish to thank them for their genuine interest in my development.

By allowing me freedom to explore ideas, Charles Oppenheim has been the best supervisor one could hope for. I am grateful for his support. I also thank members of the Information Science community for their interest. Peter Ingwersen, Christine Borgman, and Amanda Spink were especially helpful. I wish to acknowledge John Sweller and his fellow researchers within the field of educational psychology for inspiring me. John's concept of a Cognitive Load Theory is central to this research.

Special thanks go to all those who volunteered their time to participate in experimentation. Your enthusiasm was appreciated. Finally, I wish to express my deepest gratitude to the Arts and Humanities Research Board (AHRB) for awarding me a research studentship.

DEDICATION

This thesis is dedicated to Laurence, Soheila, Elisa, and Melissa.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1.1 PREAMBLE

Behavioural explanations elicited from individuals are burdened by cognitive load. Therefore within this thesis, the intention was not to set up a situation where experiment participants were required to explain their information interactions. The model of user relevance feedback interaction, developed within this thesis, has been derived from a quantitative investigation of ideas developed through qualitative research. The aim of this thesis is to show how the burden of cognitive load can be quantified so the effect it has on the interaction with information objects can be predicted. It cannot be claimed that this model provides a theoretical worldview enabling a full understanding of relevance feedback interaction.

The development of information science has been hindered in the past by theories that attempt to represent multivariate behavioural phenomena using closed theorems. The nature of this research ensures that it can be modified and expanded upon. It is anticipated that the explanatory and predictive power of the proposed model will improve as more studies of information behaviour are incorporated. Where possible, this research has encouraged experiment participants to behave spontaneously during interaction. Participants were able to complete tasks by interacting with information 'in their own way', allowing more aspects of behaviour to be investigated. This research does not attempt to explain information behaviour by simplification, categorisation, or idiosyncratic theory building. Understanding cannot exist in an artificially constructed world that is perfectly described by logical constructs. Information behaviour cannot be explained by an algorithm of general concepts. Focussing on an individual's interaction with information objects enables a connection to be made between cognitive / behavioural assumptions and real situations.

Understanding phenomena often means lifting them out of a natural habitat and inserting them into a model. One transformation may be better than another in the sense that it permits or even explains what for the other transformation remains unsolvable.

Experiments to a greater or lesser extent interfere with nature and their results are processed in 'unnatural' ways. Interference, however, has its limits. Nature is not something formless that can be changed into any shape; it resists and by its resistance reveals its properties and laws. Good experimentation eliminates disturbances, creates strong effects, and enables the underlying 'machinery' to be observed. A good model enables one to make sense of scientific results that seem to run counter to existing theories. Those who theorise about nature can be roughly classified into two groups. 'Lumpers' want to unite what 'splitters' want to separate. Lumpers can attack the arguments of splitters one by one and thus weaken the intellectual resistance to unification. Splitters argue that there is no simple scientific map of reality, but there are many different maps of reality from different scientific viewpoints. Many scientific laws, methods, and disciplines are restricted to special domains. For example, the laws of hydrodynamics are not valid in elementary particle physics even though they are often studying the same phenomena. Analysis can be stopped at various levels of specificity. One of the most fundamental twentieth-century theories, quantum theory, suggests that properties of elementary particles are not inherent but emerge as a result of special interactions. Important elements of a phenomenon are not always concealed within layers of specificity. Humans behave in complex and characteristic ways, which, though sometimes conforming to a pattern, constantly reveals surprising features. Like quantum theory, the model proposed within this thesis allows behaviour to emerge as a result of interactions, negating the necessity for 'lumping' or 'splitting'.

1.2 MOTIVATION FOR THE RESEARCH

Brookes (1977) proposed a 'fundamental equation' that attempted to model the relationship between information and knowledge.

$$K[S] + \Delta I = K[S + \Delta S]$$

where

$K[S]$ = Knowledge structure

ΔI = Change in the knowledge structure caused by assimilating new information

ΔS = The effect of the change

EQUATION 1.1

Equation 1.1 implies that the growth of knowledge is not simply accumulative. Assimilating new information may change the nature of existing relationships between concepts. Brookes (1980) claimed that his equation applies to both subjective and objective knowledge. He argued that the publicly observable growth of knowledge as recorded in published literature reflects the ways in which individual minds think privately. However, viewing knowledge as a structure of concepts linked by their relations, and information as a small part of that structure, is too simplistic.

Otto Neurath's boat, 1932:
We are like sailors who have to rebuild their ship on the open sea, without ever being able to dismantle it in dry-dock and reconstruct it from the best components (Cartwright et al., 1994).

Assimilating new information causes a transfiguration of an individual's knowledge. Subjective knowledge emerges as a result of this transfiguration process. Subjective knowledge cannot be preserved because it cannot be *dismantled in dry-dock.*

Gottfried von Leibniz, 1703:
Knowledge is determined by evidence and reasoning (Hacking, 1975).

The cognitive Information Retrieval (IR) research paradigm recognises that all interactions are mediated by knowledge states. The representation of knowledge states has considerable implications for the development of next generation IR systems.

For a system of given objects there exists alternative cognitive potentials at each given point in time. Depending on uncertainty only one or a few are uncovered. The properties of all elements may point to the actual cognitive condition of the intrinsic information need (Ingwersen, 1996, p. 41).

Wersig (1971) suggested that it is difficult to view information only as a change of an individual recipient's state of knowledge, since it is impossible to investigate a state of knowledge directly. Information is viewed by Wersig, and by many others, as a means of reducing uncertainty. Cognitive viewpoint design principles require the representation of an individual's state of knowledge about themselves, and that with which or whom they interact. The best way an IR system can represent a knowledge state is by examining interactions that reduce uncertainty. User interaction is dependent on feedback mechanisms. However, very little is known about the cognitive nature of feedback.

> The concept of feedback did not originate from theoretical roots within information science, but from IR researchers concerns with the practical problems encountered in the development and operation of IR systems. There is a need for empirical research to investigate the cognitive aspects of feedback (Spink, 1997, p. 732).

Ellis (1989) investigated the information seeking patterns of academic social scientists. Their patterns could be broken down into six characteristic stages: starting; chaining; browsing; differentiating; monitoring; and extracting. These stages constitute the principal generic features of different individual patterns. Traditional IR research has mostly involved the controlled laboratory testing of systems, while only a minority of researchers have focussed on cognitive aspects. An alternative approach is to focus on behavioural aspects of retrieval interaction, specifically, on how searchers interact with information sources. If behavioural characteristics are identified, and the retrieval system is provided with facilities that reflect those characteristics, then the users should be able to recreate their own information seeking patterns while interacting with the system (Ellis, 1989). Any facility that allows the recreation of a user's information seeking patterns is reliant on feedback mechanisms. Astoundingly, the majority of IR system designers still do not even attempt to allow users to recreate their own information seeking patterns. Those that do often find that such facilities are under-utilised (*see* Sections 2.2 & 2.4). This thesis suggests that under-utilisation is a symptom of the dissatisfaction caused by the complex, misunderstood cognitive nature of feedback mechanisms.

One method of observing information seeking behaviour is through the examination of the feedback mechanisms used during IR. User relevance feedback is one of the best known attempts at supporting / recreating user behaviour, and is a primary focus of this thesis. Other methods include browsing mechanisms, clustering, filtering, profiling, etc. User relevance feedback is an interactive technique adopted by IR systems that aims to provide the user with more control over the search process. It is achieved by re-weighting query terms based upon their distribution in documents / passages that have been judged as useful or not useful. The addition of new terms that exist in documents / passages that have been judged as being useful is also possible.

The definition provided by Chang and Rice (1993), in reference to the potential benefits of interactivity is an important insight into the cognitive nature of feedback mechanisms:

> Interactivity reduces the…"burden with respect to *cognitive load* of specifying what is needed or intended because individuals may interact directly with informational stimuli that are potentially useful" (p. 257).

Results obtained from laboratory testing suggest that user relevance feedback can significantly improve IR performance (Harman, 1992). However, outside the laboratory, user relevance feedback techniques are rarely utilised when implemented (Jansen, Spink, and Saracevic, 2000). **The main hypothesis to be tested by this thesis is that a user's ability to submit relevance feedback is burdened by cognitive load.** Limited mental resources mean that only a few elements of information may be attended to at any given time. What constitutes an element of information depends on the knowledge schemata held by an individual (Cooper, 1998).

The motivation for undertaking this research is that information science has largely failed to recognise that **an individual's cognitive limitations affect information behaviour** and dictate the way in which new information is assimilated.

Assumptions:

- Knowledge emerges as a result of a 'transfiguration process'. Information should not be viewed as a small part of a knowledge structure.
- Interactions that reduce uncertainty enable an insight into a user's state of knowledge.

Conjectures:

- Interaction is to some extent driven by cognitive load.
- A cognitive load theory of information behaviour should have both an explanatory and predictive power.

1.3 RESEARCH PROBLEM DEFINITION

The traditional approach to IR aims to maximise the retrieval performance of an IR system by using the results obtained from comparative testing. This approach is based upon the underlying assumptions of reductionism, objective observation, and linear causation. These assumptions are necessarily rigid and prevent the study of complex systems. Genuine user information seeking behaviour cannot be studied without taking into account dynamic and situational factors.

The focus of the user-orientated approach to IR is on information seeking behaviour. Schamber, Eisenberg, and Nilan (1990), amongst many others, have developed evaluation questions that aim to examine how users perceive information relative to their information need situation. Ingwersen (1992) acknowledged that investigations of both the socio-behavioural and psychological aspects of IR interaction are necessary. However, he also suggested that these types of studies could not elucidate why information processes occur on an individual mental scale. .

One important approach is the idea that the recognition of an anomaly in an individual's state of knowledge motivates that individual to use an IR system. Belkin's (1977) Anomalous State of Knowledge (ASK) hypothesis is representative of this cognitive approach to IR. The problem of uncertainty is inherent in the cognitive approach. De Mey's (1977) view of information processing is the foundation of the cognitive approach:

- **A monadic stage** - Information units are seen as self-contained entities.
- **A structural stage** - Information is seen as a complex arrangement of units.
- **A contextual stage** - Interpretation of meaning.
- **A cognitive or epistemic stage** - Information is seen as supplementary or complementary to a conceptual system.

Kochen (1983) believed that the goal of information science is to explain the nature and dynamics of information and knowledge. Logical constructs, which attempt to represent the mental state of a user, cannot convey contextual meaning. Progressing beyond the monadic and structural level requires modelling of the user's problem space and state of uncertainty.

When examining information interactions, this thesis is concerned primarily with: how an information need is developed; how it is represented; how both are affected by the technical and conceptual tools available; what problem-solving strategies are adopted; and what reasons there are for adopting these strategies. If a *sufficient* number of variables, constraints, and interactions are identified, it is theoretically possible to mathematically model any phenomenon. Unfortunately, when examining cognitive processes, it is impossible to discover when a sufficient number of properties have been found. This is because humans do not perform exhaustive searches of a problem space, even if the problem is simple enough to make it feasible. Humans use heuristics instead (Richardson, 1986; Johnson-Laird, 1983). Heuristics (or rules of thumb) direct the strategies and tactics that a user applies during a search. Heuristics impose assumptions that are used to address a problem in a way that is compatible with an individual's problem-solving abilities.

Traditional approaches to relevance feedback in IR operate within the monadic and structural stages of information processing. Relevance feedback algorithms consider how relevant terms, identified by a user, are used within documents so that the features, that might make a document relevant, can be better represented in modified queries (*see* Section 2.4). Theme and context characteristics are incorporated within most relevance feedback algorithms. A theme characteristic is calculated on the assumption that the less evenly distributed occurrences of a term are in a document, the more likely the term is to correspond to a localised discussion in the document. Context is determined by examining phrases / structures in which the term appears.

In an advanced relevance feedback system, the user is able to provide relevance feedback in a variety of ways (*see* Section 2.4). Flexibility will no doubt improve the effectiveness of relevance feedback systems. However, much relevance feedback research has used artificial pre-judged relevance judgements, with a lack of knowledge on how human relevance judgements occur and develop in an interactive IR environment. Belkin, Oddy, and Brookes (Republished, 1997) claimed that relevance feedback seems more responsive to the information need issue, in that it utilises user relevance judgements, and therefore is not solely dependent on the characteristics of the document collection. However, relevance feedback remains within the best-match (physical) paradigm, since it assumes that the eventual modified query formulation is equivalent to the ideal document.

The evaluation of relevance feedback mechanisms has only occurred at monadic and structural levels. Outside the laboratory environment, user-based relevance feedback techniques have been of limited success (Jansen, Spink, and Saracevic, 2000). IR system developers have found that user-based relevance feedback tools are unpopular. Attempts at incorporating user-based tools are often abandoned after discovering that users are frequently reluctant to provide feedback. An evaluation procedure is needed at the epistemic level to understand why relevance feedback is not effective outside the laboratory environment. The user-centred and cognitive approaches to IR attempt to operate within the contextual and epistemic stages of information processing. Without considering the extent to which potentially useful information is recognised or re-configured by a user, relevance feedback cannot convey meaning.

When evaluating relevance feedback systems using human participants, a wide range of situational and behavioural variables can be examined. Analysing the interaction between variables after a search task is completed is not realistic. This thesis will elucidate that the cognitive context in which relevancy judgements are generated cannot be revealed without an understanding of the cognitive load imposed during the search process.

1.4 COGNITIVE LOAD

Except within the field of educational psychology, the concept of cognitive load rarely extends beyond the ideas presented by Miller (1956). In Miller's famous paper "The magical number seven plus or minus two", a human's capacity for processing information was explored. It was concluded that short-term memory (working memory) has a limited retention. Cognitive load refers to the total amount of mental activity imposed on working memory at an instance in time.

The IR study by Hu, Ma, and Chau (1999) is typical of research that advocates attempts to minimise cognitive load during system design by recognising the limitations of working memory. Cognitive load was defined by Hu, Ma, and Chau as a measure of the information processing effort a user must expend to take notice of the visual stimuli contained in an interface and comprehend their significance. Two insightful IR studies have suggested that recognising the limitations of working memory may not be the only method of minimising cognitive load. Chang and Rice (1993) claimed that cognitive load could be minimised if browsing is permitted. Beaulieu (1997) argued that there is a need to consider cognitive load not just in terms of the number and presentation of search options, but more importantly to take account of the integration and interaction between them. Although both these studies point to *interaction* as being an important factor, they do not explain why. Many of the researchers who have investigated Human Computer Interaction (HCI), including those in IR, have used the term 'cognitive load' without demonstrating an understanding of Cognitive Load Theory.

Cognitive Load Theory was developed by educational psychologists in the field of instructional design, e.g., Sweller (1988; 1994); and Cooper (1990). Learning structures (schemata) are used during problem solving. Cooper (1998) explained that Cognitive Load Theory could be used to describe learning structures. Limited mental resources mean that only a few elements of information may be attended to at any given time. What constitutes an element of information depends on the schemata held by a person. A single element consisting of a single schema for an expert may be several elements consisting of several sub-schemata for a novice.

Schemata do not only provide the ability to combine many elements into a single element. They also have the capacity to incorporate the interactions between elements. Cognitive Load Theory requires the representation of both the intrinsic and extraneous aspects involved during problem-solving / learning. Within the field of instructional design, intrinsic load is induced by task difficulty, while extraneous load is induced by the nature / suitability of instruction (Brünken, Plass, and Leutner, 2002). If intrinsic load is high, and extraneous load is also high, then problem-solving may fail to occur. When intrinsic load is low, then sufficient mental resources may remain to enable problem-solving from any type of instruction, even if a high level of extraneous load is induced by the nature / suitability of instruction. Modifying the instruction presentation to a lower level of extraneous load will facilitate problem-solving only if the total load falls to a level within the bounds of mental resources.

Cognitive Load Theory for instructional design makes the assumption that cognitive load is a barrier to problem-solving performance. Developing a Cognitive Load Theory for IR that assesses the impact of cognitive load on retrieval performance would be an enormous undertaking. This thesis proposes a Cognitive Load Theory that exclusively addresses one component, i.e., the issue of relevance feedback.

For the purposes of this thesis, cognitive load is defined as the intrinsic and extraneous difficulty associated with providing relevance feedback during the search process[1].

Before applying Cognitive Load Theory to relevance feedback, some pivotal assumptions have to be made. These assumptions are consistent with those made in the field of educational psychology where intrinsic load is induced by expertise and complexity, and extraneous load is induced by learner activity (Gerjets, and Scheiter, 2002). **Intrinsic load is defined as the individual-specific difficulty associated with a retrieval task (dependent on user expertise and task complexity). Extraneous load is defined as the individual-specific difficulty associated with identifying pertinent information (dependent on learner activity).** These definitions will be discussed further in Chapter 7.

Based on the definitions stated above, the following argument will be examined by this thesis: If intrinsic load is high, and extraneous load is also high, then providing useful relevance feedback may fail to occur. When intrinsic load is low, then sufficient mental resources may remain to enable useful relevance feedback from any type of information, even if a high level of extraneous load is induced by the nature / suitability of the information presentation. Modifying the information presented to a lower level of extraneous load will facilitate relevance feedback only if the resulting total load falls to a level within the bounds of mental resources. For example, if an information visualisation interface depicts a large quantity of documents from various subject domains, a high level of extraneous cognitive load will be imposed. However, if retrieval task difficulty is low (e.g., if the user has a well defined information need) intrinsic cognitive load will be low, resulting in total cognitive load being within the bounds of mental resources. This increases the likelihood of useful relevance feedback.

---

[1] This thesis does not claim that this is the only way of considering the cognitive load concept. For example, cognitive load could be applied in a way that is more familiar to researchers in human computer interaction.

1.5 AIMS AND OBJECTIVES

The overall aim of this thesis is to develop a method that can enable the evaluation of user-based relevance feedback mechanisms at an epistemic level. This method requires the use of Cognitive Load Theory to reveal the epistemic context in which feedback judgements are made. Achieving this aim will enable the utility of relevance feedback submitted by a user to be predicted. The usefulness of being able to predict the utility of relevance feedback has pragmatic implications for developing more intuitive IR systems.

Aims:

1. Explore why users are often reluctant to provide relevance feedback.
2. Identify the role heuristics play during problem solving by examining the evolution of the cognitive load concept.
3. Demonstrate if and how an individual's cognitive architecture imposes limitations during relevance feedback.

Objectives needed to fulfil Aim 1:

A. Review existing research findings in an attempt to understand the problem.

B. Explore the problem by performing exploratory experimentation.

C. Develop a method that allows the epistemic context in which relevance feedback judgements occur to be revealed (requires Aim 2 to be fulfilled).

Objectives needed to fulfil Aim 2:

D. Review existing research findings in an attempt to understand how heuristics and Cognitive Load Theory affect problem solving.

E. Hypothesize how these findings affect information seeking behaviour.

F. Develop a method for testing hypotheses.

Objectives needed to fulfil Aim 3:

G. Perform experimentation (requires Aims 1& 2 to be fulfilled).

H. Evaluate results.

I. Draw conclusions, reveal implications and suggest improvements to method.

1.6 THESIS OUTLINE

Chapter Two introduces the ideas that have influenced the development of the cognitive viewpoint within information science. The cognitive viewpoint argues that the development of a relevance feedback system, which supports a user's changing state of knowledge, would need a model of situational usefulness that can be inferred from the user. It is concluded that deriving such a model may be possible if the relationship between a user's relevance judgements and their movement through the information seeking process is examined.

The possibility of developing a cognitive model of situational usefulness is explored in **Chapter Three**[2]. Firstly, an experimental procedure is reported that enabled the comparison of different relevance feedback capturing mechanisms. Results showed that passage feedback, a spontaneous approach to relevance feedback, can improve the effectiveness of an automatic query expansion. However although spontaneous feedback appears to allow a user to provide more situationally useful feedback, developing a cognitive model of situational usefulness requires information behaviour to be examined. For a feedback judgement to be useful, it must accurately reflect a user's knowledge state. For feedback to be evaluated, behavioural variables have to be examined in a way that provides an insight into individualistic processes. Evaluation derived directly from users' explanations of their situations and decisions is problematic. Users themselves don't always know what they want or expect.

Without the ability to quantitatively measure changes in cognitive structures, it is impossible to develop a cognitive model of a user that is capable of adapting to user needs. **Chapter Four**[3] reviews research findings that show that although uncertainty is a concept that can be used to better understand individualistic processes, it is not a suitable measure of an individual's ability to process information. The possibility that affective manifestations of uncertainty are symptomatic of limitations to cognitive capabilities needs to be considered.

---

[2] Findings presented in Chapter Three partially address Objective A (*see* Section 1.5).

[3] Findings presented in Chapter Four fulfil Objective A.

Chapter Five[4] introduces key concepts associated with information processing and problem solving. The relationship between learning and cognitive load is discussed. This chapter outlines research within the field of psychology that has been overlooked by information scientists. Many of these research findings have clear implications for the study of information behaviour.

Chapter Six[5] reviews Cognitive Load Theory research. Since humans have a limited capacity of assimilating new information, they purposefully construct meaning by selectively attending to that which connects with what they already know. The ability to assimilate information is burdened by cognitive load.

Cognitive load fluctuates throughout all learning (or problem solving) processes. **Chapter Seven**[6] details a theoretical postulate that suggests how cognitive load affects information seeking. This is achieved by bringing together research findings from information science and educational psychology. A mechanism is required that enables the degree to which a feedback judgement has been inhibited by cognitive load to be evaluated. It is postulated that when an information need progresses to a point where an individual is satisfied that it has been sufficiently addressed, it becomes easier to evaluate the usefulness associated with information objects that were used to address the information need.

Chapter Eight[7] reports on two exploratory studies. The results of these studies have been used to develop this thesis's main experimental procedure. The first study explores the fundamental premise that cognitive load is a measurable phenomenon. The second study provides an insight into the behavioural patterns associated with providing passage feedback. Additionally, the second study identifies the behavioural variables that must be incorporated within a cognitive load model of feedback interaction.

---

[4] Findings presented in Chapter Five partially address Objective D (*see* Section 1.5).

[5] Findings presented in Chapter Six fulfil Objective D.

[6] Findings presented in Chapter Seven fulfil Objective E.

[7] Findings presented in Chapter Eight fulfil Objectives B & F.

Chapter Nine[8] outlines the investigative methods and strategies used for the design of the main experimental procedure detailed within this thesis. The mechanisms developed by this thesis used to estimate the amount of cognitive load imposed on an individual, and to identify the factors causing cognitive load, do not attempt to measure changes in an individual's knowledge directly. This is an unsustainable research goal. Instead, the state of an individual's knowledge state is ascertained by exploring factors that have been identified by qualitative research as being both measurable and able to provide an insight into an individual's information behaviour.

Chapter Ten[9] describes the method used to perform the main experimental procedure. **Chapter Eleven**[10] presents results and argues that individuals are not consistent in applying the same processing style for every task they undertake. New search heuristics are generated when an individual is unable to assimilate information during the search process. Cognitive load is caused by the uncertainty that arises when an individual has to apply new heuristics, and when a mismatch in perceptions and actual experience causes existing heuristics to be abandoned. **Chapter Twelve**[11] outlines how predicting the utility of feedback judgements using Cognitive Load Theory is possible, and suggests how further improvements to the model can be made. **Chapter Thirteen** draws conclusions. It is suggested that the continued advancement of Cognitive Load Theory could provide information science with a framework that enables the examination of how an individual's cognitive architecture affects interaction with information objects.

---

[8] Findings presented in Chapter Nine partially fulfil Objective C (*see* Section 1.5).

[9] Findings presented in Chapter Ten partially fulfil Objective C.

[10] Findings presented in Chapter Eleven fulfil Objectives C & G.

[11] Findings presented in Chapter Twelve fulfil Objectives H & I.

2.0 INTRODUCTION

This chapter provides a general introduction to the ideas that have influenced the development of the cognitive viewpoint paradigm in information science. IR based on the best-match (physical) paradigm assumes that it is possible for the user to specify precisely the information need, and that the need is functionally equivalent to document texts. Relevance feedback is a significant attempt to resolve the problems inherent in the best-match principle, as a user's information need does not have to be specified precisely from the outset. However, Belkin, Oddy, and Brookes (Republished, 1997) argued that relevance feedback remains within the best-match paradigm, since it assumes that the eventual query formulation is equivalent to the ideal document. Developing a relevance feedback system using cognitive viewpoint design principles is problematic. The cognitive viewpoint argues that the development of a feedback system, which supports a user's changing state of knowledge, would need a model of situational usefulness that can be inferred from the user. It is concluded that deriving such a model may be possible if the integral relationship between a user's relevance judgements and their movement through the information seeking process is examined.

2.1 THE COGNITIVE VIEWPOINT

The desire for the efficient sharing of information has led to the development of IR technology. This technology has attempted to address the problems associated with the representation, storage, access, and searching of information required by users. IR is about the effective communication of relevant 'documents' to an individual's information needs. The notion of relevance is central to IR. The effectiveness of an IR system can be measured in terms of how long it takes for a user to find sufficient relevant information, or discover that no relevant information exists.

IR is a multi-disciplinary research field of interest to information scientists and computer scientists. Cognitive scientists and linguists have also contributed to the field. Traditionally, although not exclusively, information scientists have been more interested in behavioural and cognitive aspects of IR systems, while computer scientists have preferred to work on search algorithms. Within IR, it can be suggested that three paradigms exist that reflect these areas of interest. These are known as the physical (best-match), cognitive, and user-oriented (behavioural) paradigms.

The concept of a paradigm as a key element in the progress of science was introduced by Kuhn (republished, 1996). Ellis (1992) discussed the applicability of the paradigm concept in relation to IR, and concluded that a paradigm, in the same way as an analogy, breaks down when it is pushed too far, and this in turn produces difficulties for science based on that paradigm. Ellis believed that Cleverdon likened the Cranfield II testing environment to that of a wind tunnel (an artificial environment for a physical system). The analogy at the heart of the Cranfield tests, and which gives it its paradigmatic power, is that of the IR system being, in some way, like a physical system, both in relation to its nature and in relation to the experimental techniques appropriate for its study.

One alternative to work informed by the physical paradigm in IR research has been that which has followed the cognitive viewpoint. At the heart of the cognitive viewpoint is a retrieval system that operates using a model of the cognitive state of the user. That system must employ a representation of that cognitive state, and, in that sense, can be conceived of as having something like a cognitive identity. Ellis noted that, unlike the physical paradigm, there might be no equivalent to the Cranfield tests to serve as an analogy for those adopting the cognitive approach in this field. Although the cognitive approach has represented the main alternative focus for research, its diffuse origins and fields of application may impede the development or recognition of an analogy. Ellis concluded by suggesting that neither the physical or cognitive paradigm can be entirely separated to facilitate paradigm articulation or accelerate theory development.

The Anomalous State of Knowledge (ASK) hypothesis was proposed by Belkin (1977). The ASK hypothesis states that an information need arises from a recognised anomaly in the user's state of knowledge concerning some topic or situation and that, in general, the user is unable to specify precisely what is needed to resolve that anomaly. Thus, for the purpose of IR, it was speculated that it is more suitable to attempt to describe that ASK, than to require the user to specify their need as a request to the system.

IR based on the best-match principle (physical paradigm) assumes that it is possible for the user to specify precisely the information need, and that the need is functionally equivalent to document texts. The assumptions underlying the physical paradigm mean that such systems cannot use information from the user about uncertainty or suspicion of inadequacy in the user's state of knowledge. These, however, are the factors that prompt people to use IR systems.

The cognitive view of the IR situation (cognitive paradigm) suggests that interactions of humans with one another, with the physical world, and with themselves, are always mediated by their states of knowledge about themselves and about that with which or whom they interact. Cognitive view design principles require representation of the user's anomalies, evaluation in terms of the problem the user faces, and iteration and interaction in retrieval. Dynamic modification of search strategies and tactics, learning from retrieval mistakes and permitting modification of pre-established user models increases the usefulness of an IR system. Although the cognitive paradigm has successfully highlighted the deficiencies of the physical paradigm, it has thus far failed to provide an IR system capable of representing and translating a user's ASK.

Hert (1997) suggested that a new thread of user-oriented IR research is beginning to develop. She termed it 'naturalistic, process-oriented'. Hert suggested that researchers need to employ methods that will allow them to explore a number of possibly relevant variables and their relationships at the same time. This should take place in settings as close to real information retrieval environments as possible. The use of inductive methods enables a researcher to explore variables which might not be accessible if a particular framework drives the research. In addition, new variables may be uncovered and new theories may be developed which can inform other frameworks. Hert argued that the acceptance of dynamic processes which may be unique to a setting and context provides the possibility of achieving an in-depth understanding of that situation. The importance of a non-paradigmatic process-orientated approach to IR is probably now accepted among most researchers who are concerned with interaction.

Daniels, Brooks and Belkin (republished, 1997) proposed that an analysis of human-human information interactions reveals a general problem structure for document retrieval that could provide a means for driving human-computer interactions. They aimed to identify which individual sub-goals must be completed before appropriate query formulation is achieved. By grouping together concepts or themes identified at any given point in a dialogue, which are in the focus of attention, it is possible to identify structural elements. The amount of iteration present in the sequencing of foci suggest that the measure of goal attainment is a dynamic one. New information elicited during a particular focus may propagate through the problem structure and alter the thresholds of other goals. This technique has yet to be fully exploited, but could be integrated within many approaches to IR (*see* Section 4.2).

Proper and Bruza (1999) suggested that a knowledge discovery system should be able to derive the exact fragments of knowledge the user requires from relevant information. This would ensure that users would not need to read an entire document, and the system would give an exact and concise answer. A formal (logic-based) framework was presented where an information need is a demand for information particles, and relevance is supply meeting demand. Although this approach is interesting and well formulated, unfortunately relevance cannot be treated in such a simplistic manner. Proper and Bruza advise that a miscommunication between user and discovery system may occur, usually resulting in the selection of irrelevant information. They suggest that an information discovery system can learn from user preferences and anticipate further preferences based on those it has. They concluded by suggesting that the 'user-centred' nature of information discovery is a troublesome aspect and attempts should be made to integrate it into a formal framework. Unfortunately, it is highly unlikely that a unifying underlying theory for information discovery can be established using a logic-based framework or even a quantum-based one. The process of assigning meaning to information prevents information behaviour being explained by an algorithm of general concepts (*see* Chapter 7). This thesis focuses on an individual's interaction with information objects, enabling a connection to be made between cognitive / behavioural assumptions and real situations.

Developing an IR system that can predict information requirements is impossible as information has an individual-specific affect. Saracevic (republished, 1997) suggested that different views arise because relevance is considered at a number of different points in the process of knowledge communication. An information need is a psychological state associated with uncertainty, and with the desire to know an unknown. The idea of uncertainty initiating an information need is vital, and is discussed later (*see* Section 4.1). The concept of an information need brings about the notion of pertinence. Relevance is the property which assigns documents to a question. Saracevic suggested that pertinence is the property that assigns documents to an information need. Some other researchers take a different view, but Saracevic's definition is the one that will be used within this thesis. Some relevant answers are also pertinent; but there could be relevant answers that are not pertinent, and pertinent answers that are not relevant. Currently, IR systems can only provide relevant answers.

None of the theories of relevance are complete in the sense that they incorporate all aspects of relevance. Each theory illuminates some aspect of relevance and provides a different method for describing the properties and relationships of the notion. Saracevic identified seven universal properties of relevance. When discussing issues associated with relevance, this thesis will consider the following properties defined by Saracevic of primary importance:

1. Prior existence of a body of *knowledge.*
2. Process of *selection* concentrating on 'elements of knowledge'.
3. Selectivity based on *inference.*
4. *Mapping* of selected elements onto other 'elements of knowledge'.
5. *Dynamics*: Interaction among elements are involved; changes in any element over time is possible.
6. *Association*: The internal structure of 'elements of knowledge' affect the dynamics and vice versa.
7. *Redundancy*: More than one set of 'elements of knowledge' may satisfy the criteria.

Essentially the cognitive view assumes that a variety of individual differences in cognitive structures exist. De Mey (1977) was one of the first to theorise that the processing of information is mediated by a system of concepts which, for the information processing device, are a model of its world. Belkin et al's (republished, 1997) ASK hypothesis and Fenichel's (1981) exploration of different types of online search experiences are good examples of attempts made to explore individual cognitive differences. These types of studies revealed that IR systems need to be able to interpret the situational context in which an information request is made. This can be considered as the first cognitive barrier to successful IR.

The label-effect was discovered by Ingwersen (1982). He found that users, even with well defined knowledge of their information need, tend to label initial requests for information by means of very few concepts. This research identified a second cognitive barrier to successful IR. Traditional IR systems are not capable of distinguishing between users with detailed, some, or no knowledge about their information requirements.

Belkin, Oddy, and Brookes (Republished, 1997) argued that the cognitive viewpoint suggests that interactions of humans with one another, with the physical world and with themselves are always mediated by their states of knowledge about themselves and about that with which or whom they interact. Cognitive view design principles require representation of the user's anomalies, evaluation in terms of the problem the user faces, and iteration and interaction in retrieval. The ASK hypothesis, amongst others, has led to the development of interactive IR systems such as OKAPI (Beaulieu and Jones 1998). Such interactive IR systems attempt to support a user's changing state of knowledge. In other words, they consider how a user's state of knowledge can influence the IR process.

Ingwersen (1999) suggested that results from the first period of cognitive research (1977-1991) fell short in three ways:

- *Lack of realism.* Too much emphasis was placed on the analysis of user and human intermediary behaviour during interaction without considering embedded cognitive structures within the system and the contextual environment. Belkin & Kwasnik's (1986) study was the only significant attempt to explore which combinations of retrieval techniques were most appropriate for different information needs.

- *Lack of theory integration.* Dervin & Nilan's (1986) sense making theory was not sufficiently compared or integrated with Winograd & Flores' (1986) socio-hermeneutic approach although they both concern the same communication phenomenon.

- *Lack of holistic perspective.* Ellis (1989) questioned the ability of the cognitive approach to IR, on its own, to support empirical investigations that are different from traditional user-system studies.

A more holistic cognitive view emerged in the 1990s. The understanding of the relevance issue as a dynamic and complex phenomenon may have triggered its emergence. Robertson & Hancock-Beaulieu (1992) observed that the cognitive revolution encouraged experimentation to consider the dynamic nature of information need development and human relevance assessments. The interactive revolution has encouraged experimentation to consider situational needs and seeking processes. Ingwersen (1999) identified a further shift in focus towards a situational context that assumes consideration of the individualistic process of interpretation.

> In human information processing, the world model constitutes the cognitive space, consisting of highly dynamic autopoietic cognitive structures, that controls the perception and further processing of external input (Ingwersen, 1999, p. 14).

This thesis does not follow the traditional cognitive approach to IR. It attempts instead to explore factors that have been identified by qualitative user-based research as being both measurable and able to provide an insight into an individual's information behaviour (*see* Chapter 4). The fundamental difficulty associated with developing IR systems using the cognitive approach is one of measurement. Wersig (1971) suggested that it is difficult to view information only as a change of an individual recipient's state of knowledge, since it is impossible to investigate a state of knowledge directly. Measuring changes in an individual's 'highly dynamic autopoietic cognitive structures' is an unsustainable research goal.

Information seeking and models of behaviour are governed by a large number of variables that cannot always be measured or identified. Saracevic, Kantor, Chamis, and Trivison (republished, 1997) outlined some variables that can be measured. These fall into three main groups.

The context of an information need:

1. Problem underlying the question (perception of the problem by the user).
2. Intent for use of the information by the user.
3. Internal knowledge state of the user in respect to the problem at hand.
4. Public knowledge expectations or estimate by the user.

Strategies adopted during information seeking:

1. Observing effects of constraints on questions as indicated by users (precision, application, restrictions, limitations).
2. Describing and testing a structure of questions in IR (lead-in, query, subject).
3. Classification scheme oriented towards grouping questions by characteristics. (domain, clarity, specificity, complexity, presupposition).

Certain individual styles and decision making attributes can also be examined:

1. Ability to make inductive inferences (tested by Remote Associates Test).
2. Ability to make deductive inferences (tested by Symbolic Reasoning Test).
3. Preferred style of learning (tested by Learning Style Inventory).
4. Online experience (identified by Questionnaire).

Unfortunately, it is difficult to understand information behaviour from the examination of individual variables because they tend to interact and overlap with each other. However, this has not stopped models of behaviour being proposed using a very limited subset of variables. This problem is discussed further in Section 4.2.

2.2 USING IR SYSTEMS

Typically, a search event begins when a user submits an information request in the form of a query. Ingwersen (1992, pp. 55-56) explains that a query is a transformation of an information request, according to the logic of the actual IR techniques employed (search technology), and the information representations (documents) processed by the system at search time. Very few assumptions can be made at search time because of the quantity and diversity of users. For example, domain knowledge, the ability to make inductive inferences, the ability to make deductive inferences, preferred style of learning, and online experience will vary greatly.

A basic search query can comprise of a combination of one or more keyword(s) or phrase(s). Some IR systems can process syntactically valid question queries by using Natural Language Processing techniques or Question Template Matching. Queries can also be supplemented by the use of Boolean connectors between keywords and more advanced functionality can be achieved via the use of filters. Representation of a user information need is not a simple problem. This problem led to the development of the ASK hypothesis.

Overmeer (1999) identified that more complicated search features demand more of the user. However, he suggested that the improved retrieval performance associated with 'guided exploration techniques' will reduce overall search time. 'Guided exploration techniques' can help the user construct a query by progressively applying filters to non-specific queries using cluster visualisation. These techniques are becoming increasingly popular. Each cluster represents a set of keywords. By considering the size and distance of different clusters, assumptions can be made about the similarity associated with topic areas.

Dynamic modification of search strategies and tactics, learning from retrieval mistakes and permitting modification of pre-established user models increases the usefulness of an IR system. Savage-Knepshield and Belkin (1999) acknowledged that casual and infrequent users encounter significantly more difficulty than experienced searchers when attempting to define their information needs. Communicating with users in natural language and at the same time providing conceptual and procedural support is important. Bates (1990) argued that many users want to take advantage of automated retrieval techniques while still controlling and directing the steps of the search themselves.

Coping with different users involves consideration of IR expertise and cognitive load. Beaulieu and Jones (1998) suggested that users need a good mental model of the system. One way to achieve this is to make its functionality highly visible. However, they also argued the need to focus on search interactions rather than interface features is most important. Interaction is reliant on feedback mechanisms. Spink (1997) explained that feedback can be, and has been, viewed from different perspectives across scientific disciplines. Specifically, feedback has been a significant and central theoretical concept in the development of many cybernetic and social models. Cybernetic models view feedback as a closed loop of signalled causality underlying automatic control processes. Social models view feedback as a loop of mutual causality underlying fundamental social processes. Spink noted that despite the concern of information science with information seeking behaviour, the underlying process of feedback has not been the source of much debate or theory-building research in information science literature.

Spink concluded by arguing that previous feedback concepts were appropriate for different contexts, such as cybernetic and social processes, and that these concepts have been useful for developing models. An interactive feedback loop is a fundamental element within models of information seeking. This thesis exclusively examines the feedback process involved when a user is required to submit judgements associated with information objects. Human judgements are made using an individual-specific notion of relevance.

2.3 RELEVANCE

A data retrieval language aims to retrieve all objects that satisfy clearly defined conditions. The notion of relevance is what separates IR from data retrieval. Schamber, Eisenberg, and Nilan (1990) concluded that relevance is dependent on both internal (cognitive) and external (situational) factors. The question of defining relevance is one of determining how users perceive information relative to their information needs.

2.3.1 MEASUREMENT

User relevance judgements have been the subject of extensive research. Saracevic (1975) identified relevance as a critical factor in the emergence of information science as a discipline, and a key concept in the development of theory in the field. The universally recognised objective of an IR system is to retrieve relevant items. Some experts believe that the use of relevance as the primary evaluative criteria for IR systems has not helped the emergence of information science as a discipline, and has often hindered the development of theory. Ellis (1984) and Belkin et al (1997), amongst others, have argued that relevance should not be used as a performance criterion. Ellis argued that when using relevance as a performance criterion, the information researcher who possesses a complex understanding is unable to employ relevance as criterion suitable for measurement. In other words, the better the understanding of relevance, the less confidence that it can be employed as a measure. The continued theoretical development of this phenomenon has prevented the development of universally applicable performance metrics.

The majority of user-centred research efforts have concentrated on the examination of a user's relevance criteria, or have investigated how best to measure judgements. Relevance is not dichotomous. Judgements are made on a continuum spanning from highly relevant to partially relevant, through to irrelevant. Saracevic (1996) proposed a stratified IR interaction model. The user levels: cognitive, affective, and situational, interact with IR system levels: engineering, processing, and content, through an interface level. Saracevic defined five manifestations of relevance. Algorithmic relevance is the computed relationship between a query and information objects. Topical relevance is the relationship between the subject and information. Cognitive relevance is the relation between a user's information need and information (compatibility, novelty, quality, etc.). Situational relevance (or utility) is the relation between the task and the appropriateness of information (leading to a reduction of uncertainty). Affective relevance is the relationship between the intents and motivations of a user and the information. Saracevic admitted that some relevance levels might be hard to measure separately because they interact. Schamber (1994), amongst others, have argued that suggesting that relevance is a multidimensional phenomenon is an understatement.

Although many researchers have recognised that relevance is not a binary phenomena, when analysing results there is a tendency to categorise users' relevant and partially relevant judgements together (Spink et al, 1998). The most limited way of dealing with relevance judgements is assuming that a high level of relevance should be assigned to documents that have been judged topically relevant, independent of users. This approach is used in the Text REtrieval Conferences (TREC) where comparative testing of rival IR algorithms is performed using test collections (Harman, 1992). Strangely, the tendency to use relevance in a way that does not allow researchers to explore a user's information need is popular. The problem with using measures of quality, usefulness, and satisfaction is that they don't always correlate well with algorithmic relevance.

Harter (1992) suggested that exposure to information should have a measurable effect on the user's cognitive state. Harter argued that relevance is purely a theoretical concept. A user's information need defines their cognitive state. An information behavioural approach to relevance considers the level (cognitive, situational, affective) and context in which a user relevance judgement is made.

Spink et al (1998) explored the middle 'fuzzy' region of relevance (partial relevance). They examined if partially relevant items selected by users are related to users' level of knowledge about the problem underlying the search, changes in the users' information need, and changes in the relevance criteria employed. They found that the less users knew about the problem, the more items were assessed as partially relevant. They also suggested that partially relevant items found during an initial search may be important as they facilitate a greater understanding. Highly relevant items may not change a user's cognitive space in relation to their information problem. If more partially relevant items are identified, then uncertainty is perpetuated. Spink et al theorised that there is an integral relationship between a user's relevance judgements and their movement through the information seeking process. Based on this, they proposed that a user's relevance judgement could be plotted on three dimensions: levels of relevance (using Saracevic's (1996) stratified model), relevance region (high, partial, irrelevant), and time (i.e., the user's information seeking stage).

This approach suggests that relevance may be measurable as to its effect on the movement of a user through their information problem. For this to be possible, information seeking stages must be clearly defined (i.e., the characteristics of these stages must be clearly identifiable). Kuhlthau's (1991) stages of information seeking is a valuable insight into how stages can be identified. However, it is not realistic to treat progression through search stages as being linear, something that Kuhlthau admitted herself. Iteration of search stages makes plotting a user's relevance judgement on a 'time dimension' difficult.

2.3.2 CRITERIA

Research into the relevance judgement process clearly demonstrates that topicality (subject appropriateness / aboutness) does not automatically result in relevance for users. Models of information seeking which characterise judgements of relevance as cognitive processes based on knowledge, perception, and dynamic situational factors, have provided useful insights into the nature of relevance, e.g., Taylor (1986); Dervin (1983). Within this thesis, attempts to characterise judgements of relevance in this way will be avoided. Criteria will not be elicited directly from users as a way of evaluating the nature of relevance judgements, in contrast to, e.g., Barry (1994); Park (1993). Instead, this thesis approaches relevance from a similar perspective to Rees and Saracevic (1996). They found relevance to be a comparative rather than a qualitative concept. Various types of judgements exist because of the individual-specific interpretational purposes for which information is required.

Designing IR systems that incorporate relevance criteria other than traditional algorithmic 'aboutness' metrics is not easy. However, Barry (1994) claimed that although it may not be possible to translate all of the complex interactions that influence judgements of relevance into an IR system, it may be possible to incorporate clues that users can employ to detect qualities other than topical appropriateness. Although two users might judge the meaning of a document differently, are both judgements derived from characteristics of that document? Many characteristics of documents are dependent on an individual's interpretation. Some characteristics may well be universally identifiable, but many characteristics are surely inseparable from the process of individual-specific interpretation. Barry noted that more research is needed to examine the evaluation of information among users with varying degrees of knowledge to enable generalizable results. Domain knowledge affects the interpretation of meaning. However, there are also many other important factors, such as how cognitive load affects schema induction (interpretation), that need to be considered (*See* Chapter 5).

2.4 RELEVANCE FEEDBACK TECHNIQUES

A query can retrieve hundreds of thousands of results; it would be unrealistic for a user to try to evaluate all the results returned for relevancy. Most IR systems use relevancy ranking techniques that sort retrieved documents so that those most likely to be relevant are shown first, thereby minimising user effort. The determination of document relevancy is fundamental to the success of the information search undertaken. It is performed by calculating the frequency a query (*Q*) is matched in the index of the service provider using a combination of both established and proprietary techniques. These techniques allow the assignment of a similarity score between a document (*D*) and a query (*Q*), *similarity*(*D,Q*). Baeza-Yates and Ribeiro-Neto (1999, pp. 380-382) suggest that most IR systems currently use variations of the Vector Model to calculate similarity. A good explanation of this model can be found in Berry and Browne (1999, pp. 31-45).

In conjunction with a similarity score, Sullivan (1999, January) identified additional factors that are used to rank a Web document, such as link popularity, authoritative site reviews, and meta tags. Techniques that take into consideration hyperlinks contained within a document are becoming increasingly popular. Such techniques apply a weighting to each link based upon surrounding textual content and the similarity score associated with the linked document.

Traditionally, three proven approaches have been adopted by IR systems: Boolean, Vector and Probabilistic. These approaches work well when a user has provided a specific query. However, if a user has a less specific information need, exploration of the problem space needs to be performed (browsing). Successful exploration enables domain knowledge to be captured. This domain knowledge can then be used to provide a more specific query. Good IR systems attempt to support a user's changing state of domain knowledge by incorporating relevance feedback. However, within the Web environment and other heterogeneous domains, user relevance feedback mechanisms have not been popular with users (Jansen, Spink, and Saracevic, 2000).

User relevance feedback is an interactive process in which individuals are encouraged to utilise their domain knowledge to allow the generation of more comprehensive queries. A number of Web search engines have attempted to incorporate relevance feedback to enable query expansion. However, Jansen, Spink and Saracevic (2000) discovered that only 5% of Excite search engine users utilise relevance feedback. This finding reflects why many IR systems currently favour other query expansion techniques.

Within a heterogeneous environment, users might need to spend considerable amounts of time reformulating their queries to accomplish effective retrieval. Baeza-Yates and Ribeiro-Neto (1999) warned that this difficulty suggests that the first query formulation should be treated as an initial (naive) attempt to retrieve relevant information. Following that, the documents initially retrieved could be examined for relevance and new improved query formulations could then be constructed in the hope of retrieving additional useful documents. Three different approaches exist and can be combined:

- *Using feedback information from the user* - Feedback can be acquired from a user interactively by allowing relevancy judgements to be made, or by analysing information seeking behaviour. This is the approach to relevance feedback that this thesis is concerned with.

- *Using information derived from the set of documents initially retrieved* - By using an automated technique such as Local Clustering, the IR system can attempt to present retrieved results in a way that may facilitate faster retrieval.

- *Using global information derived from the document collection* - A technique such as Automatic Global Analysis can enable the construction of a hierarchical directory structure.

Query expansion without user relevance feedback

Both thesaurus and concept class techniques can be used to increase the number of search terms used in a query. At search time, a thesaurus can expand the original query with related terms that are similar in meaning or usage. Unfortunately, as a thesaurus is based on precompiled synonyms, related terms are often not valid in the context of a query. A concept class is a hierarchical structure that categorises terms used into potential concepts. This theoretically enables the context of a query to be preserved providing compiled classes are appropriate to the given domain.

Infoseek (1998, May) uses an Extra Search Precision algorithm (ESP) that performs like a thesaurus:

> ESP improves the quality of search results for general queries. Infoseek research shows that the majority of its users routinely use general keyword (one or two-word) searches to find information and services. ESP is especially beneficial to these users because it automatically anticipates the services and information that will be most useful to them.

Excite (1996) uses an Intelligent Concept Extraction (ICE) system that performs like a concept class technique. ICE identifies relationships between terms used within a query enabling retrieval by concepts. If a relationship is found, then the user's query is expanded with terms found within the concept class.

The Google search engine, amongst others, presents a list of terms to a user that can be added to a query based on the local clustering of results. Local clustering avoids some of the contextual problems associated with thesaurus and concept class techniques because it is more domain specific. Concepts can be extracted from documents retrieved, by the original query, based on their co-occurrence with query terms. However, Peat and Willett (1991), amongst others, recognised that frequent terms in a document set tend to discriminate poorly between relevant and non-relevant documents. The general effect is to suggest terms that do little to improve the discriminatory power of the original query.

Instead of the use of co-occurrence data, related terms can be obtained by considering terms as concepts within a concept space. The construction of a global similarity thesaurus is possible if each term is indexed by the document in which it occurs. This allows the selection of terms (computed in a term-concept space) that might be close to the whole query but distant to individual query terms.

A global similarity thesaurus is an example of an Automatic Global Analysis (AGA) technique. Baeza-Yates and Ribeiro-Neto (1999, pp. 117-139) explained the mechanics of a variety of AGA techniques. AGA could be utilised to enable the construction of a hierarchical directory search engine similar to a manually constructed directory such as Yahoo. Directories are useful when a user has a general information need. However, this general information need has to be specific in its generality. Some information needs are more complicated and require domain knowledge to be acquired before the need can approach being fully formulated. Query expansion without relevance feedback can improve the likelihood of relevant information being retrieved. It cannot help users to translate any newly acquired domain knowledge into a more effective query.

Query Expansion using Relevance Feedback

Relevance feedback can be used to provide a revised query. This is achieved by re-weighting query terms based upon their distribution in relevant and non-relevant documents that have been identified by the user. The addition of new terms that exist in documents which have been recognised as being relevant is also possible. In 1965, Rocchio (republished, 1971) performed experiments that involved the use of vector representations of both terms within a query and terms within a document to allow the relevance feedback process to take place. This developed into the Vector Space Model. Robertson and Sparck Jones, amongst others, further developed the relevance feedback concept within the probabilistic model, which is very much at the heart of IR research today (Sparck Jones, Walker, and Robertson 1998).

According to Salton and Buckley (republished, 1997) the relevance feedback process is easy to use and can prove effective. The principal idea involves the selection of important terms, or expressions, attached to certain previously retrieved documents that have been identified as relevant by the user, and of then enhancing the importance of these terms in a new query formulation. Analogously, terms included in previously retrieved non-relevant documents can be de-emphasised in any future query formulation.

Three common characteristics of relevance feedback are:

- It shields the user from details of the query formulation process, and permits construction of useful search statements without intimate knowledge of the collection make-up and search environment. Much debate exists on whether this is an advantage or a disadvantage.

- It breaks down the search operation into a sequence of small search steps, designed to approach the wanted subject area gradually.

- It provides a controlled (automatic or manual) query alteration process designed to emphasise some terms and to de-emphasise others, as required in particular search environments.

During query modification, Salton and Buckley demonstrated that a simple vector modification process that adds new query terms and modifies the weight of existing terms appears to be the most useful. Robertson (1990) argued that different weighting formulae are required for query expansion and term weighting. He used ideas from the Swets (1963) model of IR system performance. The Swets model examines the distribution of match function values over the document collection. If the retrieval system is effective, then the distribution of non-relevant documents and relevant documents will be different, i.e., the match functions will generally be higher for relevant documents than for non-relevant. The more the two distributions are separated, the better the performance of the system. When considering the process of term selection, the population of relevant documents and non-relevant documents is further divided into those that contain the new term and those that do not.

Query expansion using relevance feedback can reflect a user's fluctuating state of domain knowledge. Van Rijsbergen's (1979) association hypothesis states that if an index term is good at discriminating relevant from non-relevant documents, then also any closely associated term is also likely to be good at this. In an interactive feedback system, decisions about the association of terms should at least in part be made by the user as part of their search strategy. This allows an information request to become more precisely formulated and translated. Experimentation has shown user relevance feedback techniques are useful in the laboratory (Harman, 1992). However, as shown by Jansen, Spink, and Saracevic (2000), relevance feedback mechanisms are not popular with users. Many do not understand the purpose of the exercise, or are simply not willing to spend time providing an IR system with relevance feedback. Efthimiadis (1996) warned that although search strategy formulation has been systematically studied, it is still not well understood, and hence automation of the relevance feedback process is difficult.

Chang and Hsu (1999), amongst others, discovered that IR systems that have implemented user relevance feedback have showed very little evidence of improved search effectiveness, and that many practitioners argue instead that information visualisation is the future for IR. Information visualisation techniques aim to limit the number of results returned to a user by clustering. This can accelerate the user's relevancy determination process considerably. However, it can be argued that this activity does not support the cognitive model of a user because domain knowledge is only gathered at an abstract level. The user may as a consequence discard relevant clusters during the problem definition stage. Actual domain knowledge is essential in order to make an accurate relevancy judgement.

It is surprising that so much IR research concentrates on the development of query expansion techniques that do not involve relevance feedback. Very little research has been performed on the development of methods that capture relevancy judgments from a user in different ways. This may have compromised the popularity and usefulness of user-based relevance feedback.

At the query reformulation stage, a user has often had the opportunity to interact with information objects. A user may attempt to manually, or with assistance from the system, modify a search strategy. Two aspects that need to be considered during query expansion is the source used to provide the terms / phrases, and the method used to select the terms / phrases. IR research has shown that relevance judgements are influenced by different document representations (e.g., title, citation, abstract, full text, etc.). Schamber (1994) summarised the conclusions of research in this area. She noted that users should judge the relevance of documents by looking at the most complete representation available. The interactive method allows a user to identify useful aspects of documents. Based on these judgements, the system adjusts the weights associated with the importance of existing terms, and adds new potentially useful terms to the query. The user can also manually add / delete / weight terms themselves in order to develop their search strategy.

Efthimiadis (1996), amongst others, argued that a document should be judged as relevant to a query on the merits of the information it conveys. For relevance feedback purposes, Efthimiadis claimed that we should not compare a document with what we already know or have learned from previous documents, or with how useful we perceive it to be. As a consequence, most systems require relevance judgements on a binary scale because of the way the probability weighting is calculated. In helping users understand the difference and make relevance judgements in the above context, the question "is this the sort of thing you are looking for?" can be posed (Efthimiadis, 1992). This is a similar approach used in the OKAPI online catalogue (Walker & De Vere, 1990). Almost all relevance feedback systems currently attempt to find information based on topical relevance rather than situational usefulness (*see* Chapter 5 for notable exceptions). The development of an IR system that attempts to predict usefulness is problematic. Such a system would need a model of situational usefulness that can be inferred from the user. Passage relevance feedback is most likely to allow the context of a situational need to be inferred, as selection of individual terms does not reveal context. Before we can even begin to speculate how usefulness feedback can be interpreted by a system, a more fundamental question needs to be answered: "can individuals accurately identify usefulness clues by highlighting passages?". This question is addressed within the next chapter of this thesis.

2.5 OUTSTANDING ISSUES

- Interactive IR systems attempt to support a user's changing state of knowledge. In other words, they consider how a user's state of knowledge can influence the search process. The assumptions underlying the best-match paradigm mean that such systems cannot use information from the user about uncertainty or suspicion of inadequacy in the user's state of knowledge. Hence traditionally, for relevance feedback purposes, documents retrieved should not be compared with what we already know or have learned from previous documents, or with how useful we perceive it to be.

- User-based relevance feedback tools are not popular with users. Very little research has been performed on the development of methods that capture relevancy judgments from a user in different ways. The development of a relevance feedback system that supports a user's changing state of knowledge would need a model of situational usefulness that can be inferred from the user. Deriving such a model may be possible if the integral relationship between a user's relevance judgements and their movement through the information seeking process is examined.

3.0 INTRODUCTION

In this chapter, an experimental procedure is reported that enables the comparison of three different relevance feedback capturing mechanisms. The overall aim is to discover if the utility of relevance feedback improves when a feedback mechanism allows a user to provide 'situationally useful' feedback. Furthermore, the possibility of developing a cognitive model of situational usefulness is explored. For a feedback judgement to be useful, it must accurately reflect a user's changing knowledge state. For feedback to be evaluated, behavioural variables have to be examined in a way that provides an insight into an individualistic process. It is concluded that evaluation derived directly from user explanations of their situation is problematic. Users themselves don't always know what they want or expect.

3.1 EXPLORATORY STUDY 1: AIMS & OBJECTIVES

In the previous chapter, it was noted that there is a lack of research performed on methods that capture relevancy judgments from a user in different ways. The extent this might have compromised the usefulness and usability of relevance feedback systems needs to be established.

Aims:

1. Evaluate the usefulness and usability associated with different relevance feedback capturing mechanisms.
2. Calculate the retrieval performance improvements associated with different capturing mechanisms.

Objectives needed to fulfil Aim 1:

A. Implement a range of capturing mechanisms that enable experiment participants to submit relevancy judgements in different situational contexts.

B. Develop a method allowing the performance associated with different mechanisms to be compared.

Objectives needed to fulfil Aim 2:

C. Implement a retrieval system based on the Vector Space Model (*see* Section 2.4).

D. Implement term expansion metrics to execute revised queries enabling performance comparisons to be made within the test collection.

3.2 METHOD

39 people volunteered to be experiment participants from a pool of final year undergraduate students in the Department of Information Science (DIS) and the Department of Computer Science (DCS) at Loughborough University[12]. An e-mail had been sent to all final year undergraduates in both departments inviting them to volunteer for this study.

The first stage of the experimental procedure required the use of five search engines[13] to build a test collection. This was achieved by using information requests and relevancy judgments made by participants as a normal part of their research activity (i.e., each participant was allowed to submit a query about any subject providing it related to final year project research). Each volunteer was required to submit his or her query by e-mail (*see* Appendix A). Participants were not permitted to use advanced query modes; only keyword, phrase, or Boolean searches were permitted. All the search engines used were able to process these types of queries.

To build a test collection suitable for relevance feedback experimentation, both a participant graded relevancy score (*drs*) and a participant graded utility score (*dus*) was assigned to each document in the collection, establishing a benchmark for evaluation. For the purpose of the experiments described within this section, participants were required to assign a *drs* and a *dus* to five pages retrieved for their submitted query. Participants were able to select a score on a magnitude scale (*see* Appendix A).

Participant graded relevancy score (*drs*):
"Is this the sort of thing you are looking for? - Do not compare this page with what you already know or have learned from other pages."

Participant graded utility score (*dus*):
"How useful is this page? - Compare this page with what you already know or have learned from other pages."

---

[12] Experimentation took place February 2000. 25 students were from DIS, 14 were from DCS. There were 22 males and 17 females. All participants stated that they were experienced at using Web search engines for research purposes.

[13] AltaVista, Excite, Infoseek, Lycos, and WebCrawler.

One page returned by each of the five search engines, was selected and fed back to the participant for relevancy and utility score assignment. A selected page was not necessarily the document ranked at the top of the first results page returned by each search engine. Pages that were *undoubtedly* not topically relevant were filtered. This filtering was manually performed. The highest ranked page from each of the five search engines, that was in at least some way topically relevant, was returned to the user.

A crude relevance feedback procedure (RF_a) uses the *drs* assignment made by experiment participants to enable a revised query to be created. This new query includes extra terms extracted from documents identified as being relevant by a user. RF_a simulates the 'check-box' approach to obtaining relevance judgments. The 'check-box' approach is the most simplistic method of obtaining judgments. Documents assigned at *drs* => 0.25 were utilised for the $RF_{a0.25}$ procedure. $RF_{a0.25}$ models users who have 'checked-boxes' to provide dichotomous relevancy judgments (*see* Figure 3.1). The dichotomous approach to judgments is the most commonly adopted in IR experiments (Harman, 1992).

FIGURE 3.1: $RF_{a0.25}$

Documents assigned at *drs* => 0.75 were utilised for the $RF_{a0.75}$ procedure. $RF_{a0.75}$ models users who have 'checked-boxes' to suggest high relevance (*see* Figure 3.2).

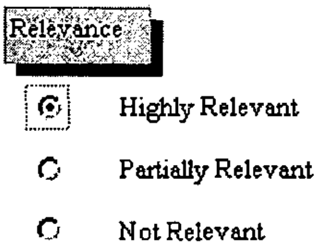


FIGURE 3.2: $RF_{a0.75}$

A more spontaneous method of obtaining relevance judgments was used for RF_b (*see* Figure 3.3). While users browsed the five documents returned, they were asked to mark what they considered to be the most useful passages using a highlighting tool (*see* Appendix A).

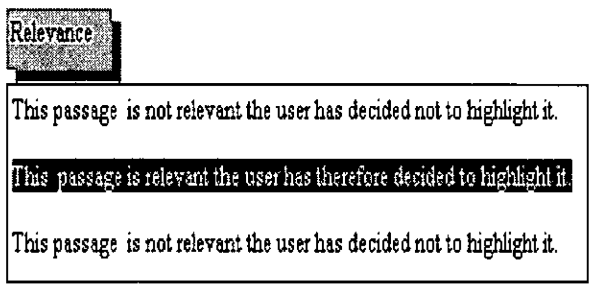


FIGURE 3.3: RF_b

The expanded queries, for both the RF_a and RF_b procedures, were created based on the following extra term selection rules. A term (*i*) within the data set identified as being relevant (*j*) was selected using a normalised frequency measure (Harman, 1992, p. 375). (*See* Equation 3.1).

$$nfreq_{ij} = \frac{Log_2(freq_{ij} + 1)}{Log_2 length_j}$$

where $freq_{ij}$ = the frequency of term *i* in data set *j*

$length_j$ = the number of unique terms in data set *j*

EQUATION 3.1

Additionally, the selected term (*i*) excluded:

- Terms that were already used in the original query.
- Terms that could be found in a stop list[14] of common words.

[14] The stop list from the SMART system located at ftp://ftp.cs.cornell.edu/pub/smart/english.stop was utilised.

The Vector Space Model, *see* Equation 3.2, was used to execute expanded queries by using the full-text of documents in the test collection (adapted from Berry and Browne, 1999). To enable performance comparisons, the Vector Space Model was also utilised as a control source for ranking without relevance feedback. DT Search[15] was used to re-rank documents within the test collection (*aj*).

$$similarity(d_j, q_k) = \frac{\sum_{i=1}^{n} (td_{ij} \times tq_{ik})}{\sqrt{\sum_{i=1}^{n} td_{ij}^2 \times \sum_{i=1}^{n} tq_{ik}^2}}$$

where td_{ij} = the i^{nth} term in the vector for document j

tq_{ik} = the i^{nth} term in the vector for query k

n = the number of unique terms in the document

EQUATION 3.2

Relevancy Ranking Evaluation

If a query expansion is successful, then the similarity score associated with more useful documents (*dus*) will be increased, enabling a more accurate ranking and faster retrieval. A method that enables the success of a ranking to be quantified is required. Participant graded utility score (*dus*) assignments were used to calculate the optimal document ranking. *Ranking Error* (*see* Equation 3.3) is the degree to which the actual ranking (*aj*) deviates from the optimal ranking (*dus*).

$$RankingError = \frac{\sum_{i=1}^{n} \sqrt{(aj - dus)^2}}{n}$$

EQUATION 3.3

15 DT Search is a powerful text retrieval tool that allows researchers and IR system developers to customise search mechanisms.

3.3 RESULTS AND DISCUSSION

Results showed that differences in relevance feedback performance were not related to sex, age, or academic discipline. Performance successes and failures were evenly distributed. The overall aim was to discover if the utility of relevance feedback improves when a feedback mechanism allows a user to provide situationally useful feedback. The RF_b procedure allows a user to provide situationally useful feedback, while $RF_{a0.75}$ and $RF_{a0.25}$ are less responsive to a user's situational needs. Figure 3.4 shows a comparison between these three relevance feedback mechanisms. The ranking error equation (*see* Equation 3.3) was used to perform this comparison. The lower the degree of error, the better the performance associated with the relevance feedback mechanism. Performance returns associated with the number terms used to expand queries are also shown. It was found that RF_b offered the best performance, while $RF_{a0.75}$ performed better than $RF_{a0.25}$.

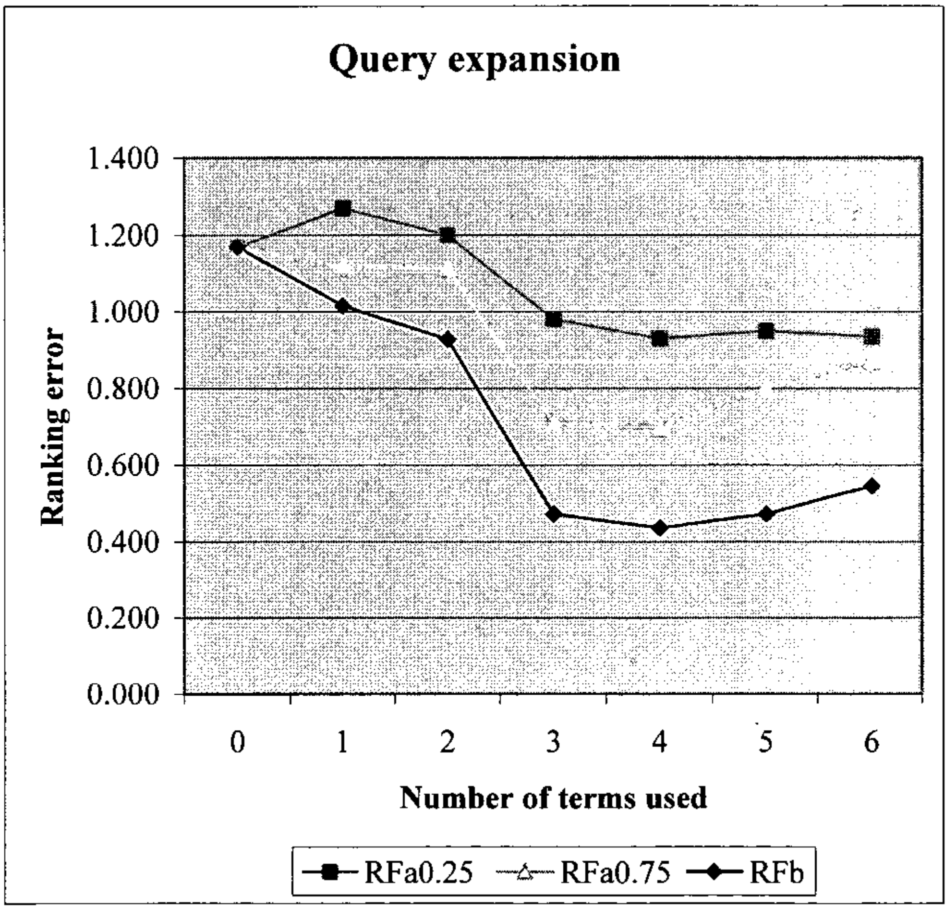


FIGURE 3.4: RANKING IMPROVEMENTS ASSOCIATED WITH DIFFERENT RELEVANCE FEEDBACK MECHANISMS

Figure 3.4 shows that relevance feedback can provide an improvement in the quality of a relevancy ranking, and as a consequence, theoretically can lead to faster retrieval of more relevant documents. This finding is consistent with the majority of studies performed within the area of relevance feedback research (Harman, 1992). However, this study is different as the procedure employed allows a comparison to be made between different capturing mechanisms. Previous research has focussed on a comparative analysis of different relevance feedback algorithms. Furthermore, many previous studies have relied on the use of artificial pre-judged relevance assignments to compare performance. This study has used a subjective measure of utility (*dus*) assigned to documents by the information requestor (i.e., the experiment participant). Therefore, the evaluation of relevance feedback is based upon the subjective utility of documents, as opposed to pre-judged topical relevance.

Almost all relevance feedback systems currently attempt to find information based on topical relevance rather than situational usefulness. Hence for relevance feedback purposes, Efthimiadis (1996) claimed that we should not judge a document on how useful we perceive it to be. Within this study, participants were asked to provide both a participant graded relevancy score (*drs*) and a participant graded utility score (*dus*). The RF_a procedure used the *drs* score for query expansion. When assigning the *drs* score, participants were asked not to compare the page with what they already knew or have learned from other pages. The RF_b procedure allowed participants to have more control over the feedback process by allowing passage feedback.

Figure 3.4 shows that $RF_{a0.75}$ unsurprisingly performs better than $RF_{a0.25}$ as the relevancy associated with a page has been specified to a greater degree. Pages identified by participants as being highly relevant were a source of more useful terms, for query expansion purposes, than terms found in partially relevant documents. Outside the laboratory environment, specifying the degree of relevancy can be complicated. A user would be required to spend a considerable amount of time cross-evaluating documents. The performance benefits associated with the labour intensive task of separating highly relevant documents from partially relevant documents may not be worthwhile.

It was found that the 'check-box' approach (RF_a) to obtaining feedback was less successful than the spontaneous 'passage feedback' approach (RF_b). It is clear that passage relevance feedback (RF_b) can improve the effectiveness of an automatic query expansion. This suggests that relevance feedback can be successfully captured within the context of a user's relevancy determination process, i.e., while browsing documents. Furthermore, it can be speculated that RF_b enables judgments that better reflect the fluctuation of both cognitive and situational factors of relevance to be captured (*see* Section 2.3). This is achieved by allowing the semantic entities of documents to be assessed for relevancy.

It can be concluded that different relevance feedback capturing methods can appreciably influence the effectiveness of a query expansion. The lack of research regarding the development of such methods has almost certainly compromised the success and popularity of relevance feedback. This experiment has enabled the comparison of relevance feedback capturing mechanisms. It can be argued that by allowing a user to provide more 'situational useful' passage feedback, the utility of relevance feedback can be improved. Supporting a user's fluctuating understanding is only possible if interaction allows for the revision of a query while a user is browsing. Spontaneous relevance feedback could be a possible solution. Implementation of a relevance feedback mechanism is not straightforward, as every interaction changes a user's state of knowledge. A view of relevance might, as a consequence also change; relevance feedback should reflect the changing needs of a user and be captured within the context of a user's relevancy determination process.

When query expansion techniques are applied to an online system, where new documents can be retrieved as a result of an expansion, the success of such techniques can be compromised. For every new term used in an expansion, there is an increased probability of irrelevant documents being retrieved. Therefore, a threshold limiting the number of terms used for expansion should be applied. For example, as shown in Figure 3.4, improvement associated with rankings stop after a four-term query expansion is used. In this case, expansions using more than four-terms are likely to be futile. Term weighting should be considered alongside threshold setting to help minimise the problem.

Relevance feedback can provide an improvement in the quality of a relevancy ranking, but outside the laboratory environment, specifying relevancy can be convoluted. A user is required to spend a considerable amount of time evaluating documents by 'checking-boxes'. RF_a requires the user to provide judgments outside the context of their browsing environment. The $RF_{a0.75}$ mechanism demonstrates that performance advantages may be possible if partial relevancy judgments are used instead of the dichotomous ones simulated by $RF_{a0.25}$. However, specifying accurate partial relevancy judgments requires time consuming evaluation iterations. The quality of a RF_a query expansion is dependent upon the quantity of systematic relevance assessments performed. Furthermore, as Moens (1999, November) suggested, Web pages will become dynamic entities, compiled from a variety of different pages (using information extraction tools), and based on user profiling implemented using machine learning techniques. Relevance feedback supplied using the 'check-box' method will not be possible within this dynamic environment. The concept of evaluating a 'stand-alone' Web page will therefore eventually become outdated.

Unfortunately, this exploratory study was only partially successful in achieving its aims. Although Aim 2 was fulfilled[16], Aim 1 was only partially addressed. **Evaluating the usability associated with feedback capturing mechanisms cannot be performed using the metrics developed for this exploratory experiment. This issue is discussed in detail below.**

Both the cognitive and situational factors affecting relevance are continually fluctuating. Ideally, a relevance feedback mechanism should be designed in a way that enables spontaneous relevance judgments to be captured. If individual semantic entities of a document (sections, paragraphs, sentences, etc) are used to generate a relevance judgment, the likelihood of more useful judgments being made is increased. However, the only way of evaluating the usability associated with feedback tools is to understand the individualistic behavioural context in which feedback decisions are made.

[16] Calculate the retrieval performance improvements associated with different relevance feedback capturing mechanisms.

There are numerous problems associated with IR evaluation that relies on the use relevance assessments. Harter (1996) argued that important information concerning individual subjective differences remain hidden. These differences arise from the fact that users possess different levels of experience, knowledge and understanding. Furthermore, MacCall (1998) suggested that a method for evaluating the relevance of information resources must allow judges to signal their cognitive capability during evaluation. Although Exploratory Study 1 evaluated the success of relevance feedback by using a measure of user satisfaction[17], individual subjective differences remained hidden and cognitive capability during evaluation was not signalled. Exploratory Studies 2 and 3 concentrate on such aspects by adopting a behavioural analysis (*see* Chapter 8).

In Exploratory Study 1, participants were instructed to interact in a specific way. Experiments 2 & 3 allow for flexibility in behaviour. Flexibility is required for the study of realistic user behaviour. The analysis of information seeking behaviour can be used to inform the development of IR systems. For example, Ellis (1989) investigated the information seeking patterns of a variety of academic social scientists. These characteristics constitute the principal generic features of the different individual patterns, and together provide a flexible behavioural model for information retrieval system design. Traditional IR research has mostly involved controlled laboratory testing of systems, while only a minority have focussed on cognitive aspects. An alternative approach is to focus on behavioural aspects of retrieval interaction, specifically, on how searchers interact with information sources. If behavioural characteristics are identified, and the retrieval system is provided with facilities that reflect those characteristics, then the users should be able to recreate their own information seeking patterns while interacting with the system.

Borland and Ingwersen (1997) described the ideas underlying the development of methods to enable the evaluation of interactive IR systems. Their method aims to collect cognitive as well as traditional performance data. Relevance assessments are made by users with reference to topical and situational relevance.

[17] Relevance judgements were only used for RF_a query expansion.

Borland and Ingwersen's method takes into account the dynamic nature of information needs that are assumed to develop over time for the same user, a variability that is presumed to be strongly connected to the process of relevance assessment. The criticism of the conventional methods of IR evaluation is centred on the static nature of the information need. In TREC, query modification and relevance feedback are used only to revise the initial search query, and not the information need. As Borland and Ingwersen's method involves partial as well as situational relevance assessments (from all test persons), it is possible to observe whether inclusion of terms from partial or non-relevant documents make a difference. This enables at least some of the cognitive and qualitative aspects of IR interaction to be examined.

The development of an information need is a result of a cognitive 'breakdown situation'. Since researchers cannot control the outcome of the situation during experiments, it is reasonable to test and evaluate by means of predefined indicative requests in the context of a simulated work task situation. Each searcher will interpret the task differently to simulate such a 'breakdown situation'. Relevance judgements are thus not based on the query submitted to the system, but relate to the user's requirements and mental state at the time of receiving the result for judgement. Based on a qualitative analysis of data collected, it can be determined whether users obtained any knowledge by browsing the retrieved documents that made them modify their queries. Borland and Ingwersen (1997) found that all users modified their queries based on insight gained from the information objects, associated with their own domain knowledge.

Spink and Saracevic (1997) investigated the sources and effectiveness of search terms used during on-line searching under real-life circumstances. From these findings, it is clear that the selection of search terms for a question and the construction of queries is a highly interactive process. It is iterative and not linear. It may take a number of retrieval cycles before relevance feedback becomes useful. Borland and Ingwersen's (1997) interactive evaluation method placed a limit on the number of iterations. This obviously introduces an artificial element.

Relevance feedback is a difficult task as it often involves decisions being made during the development of an information need. A poorly developed information need may lead to inaccurate decisions being made. This problem is amplified if an IR system performs automatic functions that are hidden from users. Belkin et al (2000) noted that previous investigations have revealed that users generally prefer to have some measure of control on what the system does for them and to understand why. They discovered, as might be expected, that users preferred the system that they perceived to be more effective. User perceptions, however, were not related to objective performance measures. Possible reasons for this anomaly include the complexity of interface, or the satisfaction associated with 'being in control' not being enough to overcome effort.

Beaulieu (1997) examined the Human-Computer Interaction (HCI) aspects of relevance feedback using the Okapi text retrieval system. Three experiments implementing different approaches to query expansion were described, highlighting the close relationships between the system's functionality and different interface designs. Key issues included the level of visibility of the system's operations, the distribution of control and the user, and the extent of cognitive load on the user. Beaulieu speculated that cognitive load appears to be dependent on two interrelated features: complexity and control. It is generally recognised that in the design of IR systems, there is a trade-off between functionality and ease of use. Beaulieu warned that in IR systems, there are different levels of functionality, and there is a need to consider usability not just in terms of the number and presentation of options, but more importantly to take account of the integration and interaction between them.

I^3R is a retrieval system that is dependent on relevance feedback (Thompson and Croft, 1989). Browsing begins after the user has selected information objects as a starting point. The system suggests paths that should lead to relevant information. Recommendations do not restrict options; neighbourhood context maps are constructed using nodes representing different linked concepts. At the centre of the concept map are the information objects selected as being relevant. The surrounding nodes are information objects that the system has identified as being potentially relevant on the basis of feedback. The evaluation of user relevance feedback when using advanced IR systems such as I^3R is not possible using traditional performance data. A model of the interactive process is needed.

Spink and Losee (1996) believed that in theoretical models of the interactive process, such as Ingwersen, 1996; Saracevic, 1996, feedback is only treated as an implied process within interactive models and is not fully elaborated. Spink and Losee argued that the challenge for interactive models is to incorporate and elaborate feedback processes and user judgements related to a variety of situational and cognitive user states, including query reformulation.

The traditional IR concept of relevance feedback is limited to topical relevance. The feedback concept clearly needs to extend far beyond this in order to allow the evaluation of interactive IR systems. **Feedback involves complex behavioural processes. Additionally, the ability to provide effective feedback is restricted by an individual's cognitive capabilities.** Interactive systems are reliant on feedback. This thesis argues that feedback mechanisms should be designed in a way that allows cognitive limitations that affect interaction during the provision of feedback to be identified. This identification process has to involve the analysis of information behaviour. In order to understand feedback, an individual's information behaviour has to be understood in the context in which it occurs. For feedback to be evaluated, behavioural variables have to be examined in a way that provides an insight into an individualistic process.

3.4 CAN RELEVANCE JUDGEMENTS BE CAPTURED?

The question of defining meaning is one of determining how users perceive information relative to their information needs. Saracevic (republished, 1997) suggested that different views of relevance arise because relevance is considered at a number of different points in the process of knowledge communication. A good relevance feedback mechanism should be able to capture spontaneous relevance judgments because both cognitive and situational factors are continually fluctuating. The ASK hypothesis (Belkin, Oddy, and Brookes, republished, 1997) rejects the possibility of an accurate dichotomous relevance judgment being made unless the ASK is resolved. Most relevancy judgments, therefore, can only be considered as being partial.

Ingwersen (1996) suggested that a second dimension of partial relevance arises. Individual semantic entities of a document (sections, paragraphs, sentences, etc.) are used to generate a relevance judgment. These semantic entities are not identical to an entire document. A good relevance feedback mechanism should be able to deal with partial relevancy judgments associated with individual semantic entities.

Measuring partial relevancy judgments can be problematic. Tang, Shaw and Vevea (1999) hypothesised that confidence in the accuracy of a relevance judgment is a function of the number of relevance categories that are available in the rating scale. They recommend a seven-point scale, as this leads to the highest level of confidence amongst users. Eisenberg (1988) promoted the use of a technique known as magnitude estimation because it is less prone to biases than commonly used category rating scales. Magnitude estimation is an open-ended scaling method requiring a user to directly estimate relevancy using a number that seems appropriate. After converting raw scores to logarithms, the geometric mean can be calculated, allowing comparisons to be made. Magnitude scales do not seem to be biased by the order of document presentation in the same way category rating scales are. Unfortunately, the judgment of a particular document is affected by the other documents being judged, regardless of the scaling method used. The geometric mean should be the same for a restricted range of documents in relation to the full range of documents for the same query. Eisenberg found that geometric means were different, indicating that the context in which a judgment is made causes bias. This provides further evidence that a good relevance feedback mechanism should allow semantic entities within documents to be assessed for relevancy.

Spink and Greisdorf (2001) examined the characteristics of the distribution of user relevance feedback judgements. A variety of approaches have been proposed that attempt to establish the best method for measuring user relevance feedback judgements. The results of Spink and Greisdorf's study provide evidence that choice of scale should not be the prime concern in developing an appropriate methodology for quantifying relevance judgements. Users appear to be quite capable of discerning differences in items retrieved from IR systems; and scales that allow a broad range of choices to accommodate those differences can be used to log relevance judgments about those differences. The capturing of user relevance feedback judgements is not simply a problem of selecting an appropriate measurement scale.

Bates (1996) found that a user often identifies topically relevant documents as being highly relevant even if they are not useful in developing domain knowledge. In fact, users are often looking for novel or unfamiliar items. Spink, Greisdorf, and Bateman (1998) suggest that a relationship between moderately relevant items and the development of a user information need during information seeking may exist. The retrieval of moderately relevant documents may have a crucial role in providing users with new information and directions, especially during the early stages of the search process. It is doubtful that a relevance feedback system could be designed to retrieve these, potentially more useful, moderately relevant documents. If a relevance feedback mechanism allows a user to be able to signal 'the degree' to which a document has transformed their state of knowledge, would this allow the cognitive context of a relevance judgment to be revealed? How could the cognitive context be used by an IR system? As previously discussed, according to Wersig (1971), examining a state of knowledge directly is not possible. The emergence of meaning cannot be predicted from perceptual elements. This is discussed further in Chapter 7. Examining interactions that reduce uncertainty is the only realistic way of modelling a knowledge state.

Wilson et al (2000) suggested that relevance could be perceived as a surrogate for uncertainty. Information can be perceived as relevant if there is a possibility that it will reduce uncertainty. How uncertainty relates to relevance and other information seeking behaviour aspects is not well researched. During experimentation, Wilson et al found that participants were able to identify which problem-solving stage they were engaged in (identification / definition / resolution / presentation). The majority of participants were able to use scales to identify their state of uncertainty. Results showed that differences in the level of uncertainty experienced by individuals are not related to sex, age, or discipline. As participant knowledge of the field increased, so did the certainty they expressed. The relationship between the level of knowledge and uncertainty about the probable availability of relevant information was not significant. Two different ideas of uncertainty were discovered: uncertainty associated with affective dimensions, and cognitive uncertainty associated with rational judgements about the problem stages.

Kuhlthau (1993) explained that uncertainty is a state that commonly causes affective symptoms of anxiety and lack of confidence. Uncertainty due to a lack of understanding, a gap in meaning, or a limited construct, initiates the process of information seeking. Although serving as a useful insight, affective symptoms do not elucidate how information processes occur.

Ingwersen (1996) developed a polyrepresentation concept that enables the representation of a user's current information need, problem state, and domain work task. Uncertainty and unpredictability are associated with acts of interpretation. Communicated messages remain at a linguistic level until they transform a human cognitive state by turning into information. Ingwersen argued that relevance discussions will continue in vain if one single definition is attempted; what is necessary is to understand which types of relevance are applied in an experiment (topicality, situational, preference, task, etc.) The very nature of IR is characterised by uncertainty and unpredictability, and these problems should not be eliminated from simulated laboratory experiments. For a system of given objects, there exists alternative cognitive potentials at each given point in time. Depending on uncertainty, only one or a few are uncovered. The properties of all elements may point to the actual cognitive condition of the intrinsic information need.

Saracevic & Kantor (1988) performed a large-scale investigation into the factors that increased the likelihood that retrieved items would be judged relevant. They found that users with a well-defined problem, users with a high estimate that information could be found, individuals who are good with word association, and individuals who prefer learning in abstract terms were likely to judge more items as relevant. This study did not explore whether the relevance judgements instigated by these factors were useful to the IR process. For a relevance judgement to be useful, it must accurately reflect a user's knowledge state. The concept of 'judgement accuracy' has not been researched by information scientists.

Saracevic and Kantor also found that when requests are low in clarity or specificity, when needs are high in complexity, and when a search is highly iterative, relevancy judgements are affected. These factors are associated with the resolution of uncertainty during information seeking. The way in which uncertainty affects 'judgement accuracy' also needs to be investigated.

Robertson & Hancock-Beaulieu (1992) speculated whether query reformulations are due to genuine changes in the searcher's state of knowledge, or due to the searcher expressing a previously unexpressed information need. They also considered whether reformulations are due to positive effects caused by user-system interaction, or just the user adapting to the system, or whether it is a combination of factors. IR interaction processes should not be bounded at points of requests and associated with information objects, but should extend into the mental states of the user. A comprehensive understanding of a user's mental state is, however, impossible to achieve. Various researchers have shown that user criteria can be implemented by IR systems by analysing the behaviour of specific user groups (e.g., Ellis, 1988). Schamber (1994) admited that due to fact that information scientists lack a full understanding of what users want or expect, the risk of making predictions on false cognitive assumptions cannot be eliminated. Users themselves don't always know what they want or expect. A goal of this thesis is to explore the extent that this type of uncertainty is a barrier to 'judgement accuracy'.

Schamber (1994) noted that the studies performed by Barry (1993), Cool et al. (1992), Schamber (1991), and Thomas (1993), were generally inspired by the idea that relevance judgements alone cannot convey the multiple meanings underlying users' decisions on whether to pursue information. The studies were qualitative, exploratory, and descriptive; criteria were derived directly from users' explanations of their situations and decisions. This thesis proposes an approach than does not rely on user explanations. Instead, this thesis will attempt to show how these explanations can be affected by the cognitive load imposed on an individual.

3.5 OUTSTANDING ISSUES

- Different relevance feedback capturing methods can appreciably influence the effectiveness of a query expansion. The lack of research regarding the development of such methods has compromised the success and popularity of relevance feedback. Results from experimentation reported within this chapter demonstrate that passage relevance feedback, a spontaneous approach, can improve the effectiveness of an automatic query expansion.

- Exploratory Study 1 was only partially successful in achieving its aims. Evaluating the usability, associated with feedback capturing mechanisms, cannot be performed using the metrics developed for this exploratory experiment, as individual subjective differences remain hidden. A method for evaluating relevance judgements assigned to information objects must allow judges to signal their cognitive capability during evaluation.

- Relevance feedback is recognised as one of the most difficult tasks that users of information retrieval systems are asked to perform. Evaluation is problematic due to the complexity of the task. To understand feedback, an individual's information behaviour has to be understood in the context in which it occurs. For feedback to be evaluated, behavioural variables have to be examined in a way that provides an insight into an individualistic process.

- Relevance judgements alone cannot convey the multiple meanings underlying users' decisions on whether to pursue information. The capturing of user relevance feedback judgements is not simply a problem of selecting an appropriate measurement scale. For a relevance judgement to be useful, it must accurately reflect a user's knowledge state. The concept of 'judgement accuracy' has not been researched by information scientists. This thesis will attempt to address this issue.

4.0 INTRODUCTION

This chapter reviews research findings that show that although uncertainty is a concept that can be used to better understand individualistic processes, it is not a suitable measure of an individual's ability to process information. The possibility that affective manifestations of uncertainty are symptomatic of limitations to cognitive capabilities needs to be considered. Without the ability to quantitatively measure changes in cognitive structures, it is impossible to develop a cognitive model of a user that is capable of predicting user needs. This is why information behaviour must be studied.

Spontaneous relevancy judgements occur during the process of browsing information objects. Browsing is often seen as an exploratory way of satisfying an information need, while searching is considered to be more goal-orientated. For example, Janes (1989) found that an adaptive search plan, which uses information gained during the search process (browsing), outperforms a non-adaptive plan (searching). There is now recognition that browsing and searching should not be treated as separate activities. Bates (1989) argued that browsing is integrated into many forms of searching. The lack of established labels to categorise various forms or degrees of browsing has prevented their elaboration. Chang & Rice (1993) believed that browsing should be considered as an important heuristic search strategy to be used in situations where the user has an information need that is low on specificity. This could be due to an inability to specify search requirements, or an unfamiliarity with the subject domain.

Browsing has been associated with visual recognition and logical reasoning (Oddy and Balakrishnan, 1991). Cove and Walsh (1988) argued that browsing is the art of not knowing what one wants until one finds it, implying that recognition is an important aspect. Chang & Rice (1993) believed that browsing is an iterative scanning and examining activity. Thompson and Croft (1989) suggested that browsing is a feedback processes that is incremental and it is the user who determines what and the depth to which information objects are examined. The type of strategy applied by individuals when browsing can be considered as different types of information behaviour. Investigation of the factors that affect information behaviour during the process of submitting spontaneous feedback judgements need to be investigated.

Marchionini (1987) suggested that one of the reasons why users browse is to reduce cognitive load. Marchionini believed that less cognitive load is imposed during browsing than it does to plan and conduct a search. Foss (1989) suggested that browsing can also result in information overload (cognitive overload). Foss argued that overload stems from inexperience with learning from browsing, which leads to difficulties when attempting to understand the information content. Additionally, Iselin (1989) believed that the complexity of information also causes cognitive overload. Foss (1989) suggested that the use of a summary text box[18] that enables users to make notes during hypertext browsing is a useful technique when attempting to deal with overload. Examining the affect browsing has on cognitive load cannot be achieved without quantifying changes in knowledge.

Ellis (1996) believed that the desire to simplify measurement as opposed to maintaining the complexity of phenomena has led to the pursuit of quantification at the expense of validity both inside and outside the experimental environment. Any quantitative investigation of changes in knowledge should be derived through ideas developed through qualitative research. Wilson et al (2000) found that the majority of their experiment participants were able to use scales to quantify their state of uncertainty. In this thesis, the concept of uncertainty is used to help explain an individual's knowledge state.

4.1 UNCERTAINTY

The concept of *uncertainty*, within the context of information processing, was first proposed by Shannon and Weaver (1949):

> Uncertainty is the critical link between information and decision-making. To effect a meaningful analysis of pragmatic information, one must look in detail at that which makes decision-making such a challenging and often agonizing activity: uncertainty (p. 224).

[18] This technique is implemented within experimentation carried out within this thesis to help with the exploration of cognitive load (*see* Chapters 9-11).

Although Shannon and Weaver's uncertainty concept was insightful, their proposed definition was too narrow[19]. Decision-making is influenced by information seeking behaviour, which means that any information type, irrespective to how pragmatic it is, has the potential to affect behaviour. Whittemore and Yovits (1973) proposed that the suitability of information for a user's decision-making could be evaluated by measuring uncertainty. Bates (1986) suggested that uncertainty could be applied as a way of accounting for variations in human thought when designing indexing systems. Both these studies make the assumption that uncertainty can be used as an evaluative metric. Whittemore and Yovits are amongst many others who have argued that good information reduces uncertainty. Bates argued that uncertainty could measure an individual's ability to process information. This thesis will avoid making such assumptions and will elucidate why making such assumptions are dangerous.

Yovits and Foulk (1985) noted that as an individual's understanding changes over time, and consequently an individual's attitude towards uncertainty also changes. They found that in the majority of situations, information reduces uncertainty. They could not provide an explanation of why, on some occasions, uncertainty increased, other than speculating that an individual found the information surprising.

Kuhlthau's (1993a; 1993b) view of the uncertainty concept is currently one of the best developed. Kuhlthau's 'uncertainty principle' is based upon a constructivist view of learning. She criticised all previous studies on uncertainty by arguing that information science should cover all levels of decision-making. For example, Yovits and Foulk's (1985) study of decision-making was limited to those who are assumed to have reached an advanced state of knowledge. Kuhlthau argued that attention should be paid to novel situations where a sequence of related decisions are required over a period of time as well as advanced decision situations:

> Research into a user's actual experiences in the process of using information for seeking meaning... reveals pervasive patterns of uncertainty. An uncertainty principle is proposed as a basic premise in the process of learning from information access and use (Kuhlthau, 1993, p. 110).

[19] Shannon and Weaver were primarily concerned with telecommunications so it is not surprising their definition is too narrow.

Kuhlthau used Kelly's (1963) 'personal construct theory' as a prototype for theory building. A fundamental postulate is elaborated by a series of corollaries. Although these corollaries explain the postulate, they are not structured or presented in any order. Kuhlthau supports her 'uncertainty principle' using six corollaries.

4.1.1 KUHLTHAU'S FUNDAMENTAL POSTULATE

The affective symptoms of uncertainty, confusion, and frustration are associated with vague, unclear thoughts about a topic or question. As knowledge states shift to more clearly focussed thoughts, a parallel shift occurs in feelings of increased confidence. Uncertainty due to a lack of understanding, a gap in meaning, or a limited construct initiates the process of information seeking.

4.1.2 PROCESS COROLLARY

Kuhlthau claimed that Dewey's (1933) 'phases of reflective experience', Kelly's (1963) 'phases of construction', and Bruner's (1973) 'interpretive task', provide a theoretical basis for understanding individual experience in using information. This thesis argues that although these theories can be used to *classify* individual experience, they fail to provide an *understanding* of individual experience. Studying a user's perspective of a search task will obviously provide a useful insight into the search process. Collating results from a large number of users reveal that individuals commonly experience a series of phases or stages as they seek information. This only provides an abstract view of the search process. It does not increase our understanding of how an individual progresses from one search stage to another. An evaluation of each individual's epistemic processes is needed to *understand* an individual's experience.

Kuhlthau suggested that during the stages of the search process, individuals construct their own understanding. She believed that affective symptoms of uncertainty (e.g., confusion/frustration) are less pronounced because an increase in understanding, interest, and confidence occurs as the search progresses to a more advanced stage. Experimentation performed by Wilson et al (2000) shows that this is not always true (*see* Section 9.5).

4.1.3 FORMULATION COROLLARY

The formulation of a focus or a guiding idea is the pivotal point in a search where a general topic becomes clearer and a particular perspective is formed as the user moves from uncertainty to understanding. This involves assimilating information, not merely locating it. Kuhlthau found that disconcerted feelings are commonly associated with idea or focus formulation. Users found the period proceeding formulation of a focus the most difficult stage of the search process. Kuhlthau argued that we have no way of knowing how many explorations have been abandoned at this point, but claimed that the shift in increased confidence parallels increased clarity as formulation unfolds. Kuhlthau viewed the formulation of a focus as an elaborative choice. Kelly (1963) suggested that elaborative choices trigger formulation and typically occur after considerable exploration has taken place. However, it can be argued that without attempting to estimate the number of explorations (in some way), and explore the nature of these explorations, the very existence of a formulation corollary will always remain a purely theoretical concept.

4.1.4 REDUNDANCY COROLLARY

Redundant information fits in with what an individual already knows and is promptly recognised as being relevant or irrelevant. Kuhlthau suggested that redundancy might be expected to increase as uncertainty decreases, and the lack of redundancy at the beginning of the search process may be an underlying cause of anxiety related to uncertainty. Unique information does not match an individual's former constructs and hence prompts an individual to construct new ideas and learn new concepts. The balance of redundancy and uniqueness is critical in the search for information. Too much redundancy results in boredom; too much uniqueness causes anxiety.

4.1.5 MOOD COROLLARY

Kelly (1963) claimed that two moods exist during any constructive process; one is invitational and the other indicative. Kuhlthau claimed that an invitational mood leads to expansive actions, while an indicative mood leads to conclusive actions. She speculated that an invitational mood might be more appropriate for the early stages of the search, and an indicative mood more appropriate for the latter stages. Kuhlthau found that case study participants who were too indicative tended to choose topics without sufficient preliminary investigation, while those with invitational moods found it difficult to select topics.

4.1.6 PREDICTION COROLLARY

Kelly (1963) theorised that individuals form constructs with which to predict future events. Kuhlthau believed that feelings of anxiety at the beginning of a search affect the choices that one makes as do feelings of confidence in later stages, and that predictions are formed through the holistic experience of information seeking. Kuhlthau borrowed Kelly's concept of using unstructured corollaries to support her fundamental postulate. If expectations/predictions are met, then uncertainty will decrease, but this may in turn lead to a potentially less useful 'indicative mood'. An 'uncertainty principle' for information science should not deny the existence of these types of interactions. The use of corollaries prevents the study of how different aspects of uncertainty interact.

4.1.7 INTEREST COROLLARY

Motivation and intellectual engagement intensify along with the construction of novel postulates. Kuhlthau found that personal interest increases as uncertainty decreases. Dewey (1933) suggested that intrinsic motivation directs the individual's action towards a satisfying conclusion. Fransson (1984) found that motivation plays a significant part in whether a person adopts a surface learning approach or a deep approach. It is much harder to adopt a deep learning approach if interest is low.

4.1.8 MANIFESTATIONS OF UNCERTAINTY

Kuhlthau claimed that her studies indicated that the process information seeking is not purely cognitive. She argued that an understanding of affective aspects of the search process could help user's learn to tolerate anxiety, i.e., when too much unique information is found. She believes that a misreading of anxiety feelings as a signal of failure is likely to occur when users do not have an understanding of the affective component of the constructive process. Kuhlthau's 'uncertainty principle' has served as a valuable insight into the nature of uncertainty. Experimentation carried out within this thesis (*see* Chapters 9-11) explores Kuhlthau's ideas by attempting to incorporate affective aspects of uncertainty within a model of information behaviour. The aim is to explore the degree manifestations of uncertainty influence information behaviour. The possibility that manifestations are symptomatic of limitations to cognitive capabilities is also considered.

4.2 INFORMATION BEHAVIOUR RESEARCH

Daniels (1986) claimed that when developing user models, crude over-generalisations might arise when the system is forced to match a user with a preconceived behavioural template. Belkin (1978) noted that the basic problem is reconciling the need for prediction with the seemingly individual-specific effect of information.

A number of models have been proposed, which can be used to explain user searching behaviour, that are of importance to the relevance feedback process. Ingwersen (1996) presents a model in which information objects, the user, and the system are all perceived as cognitive structures. Belkin's (1995) episode model involves a classification of information-seeking strategies that are used during search events. Saracevic (1996) believed that the understanding of interaction requires appreciation of the interplays between several connected strata. Although information-searching models are concerned with how searchers interact with systems, each of them represents user interaction at different levels of abstraction.

Generally, models of information behaviour do not attempt to describe the same types of activities. Information behaviour can be defined as the activities of an individual during the identification of needs, searching, using, or transferring information.

This thesis is predominantly concerned with the use of relevance feedback within IR systems. It is, however, essential to be aware of more generalised behaviours surrounding the initiation of information seeking. The excellent paper by Wilson (1999) highlighted the problems associated with viewing models of information searching, information seeking, and general information behaviour, independently. Figure 4.1 shows that these models can be viewed as complementary, rather than competing.

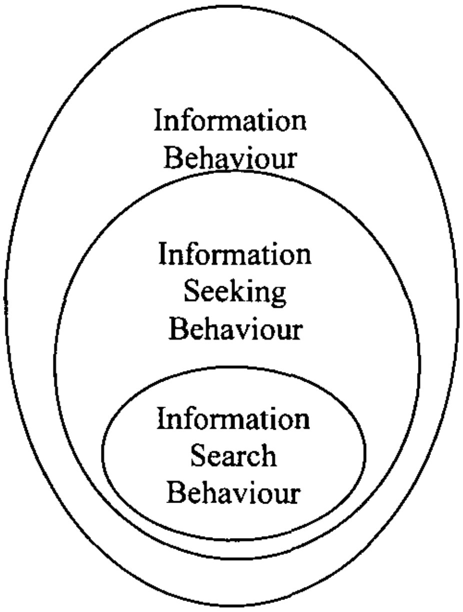


FIGURE 4.1: WILSON'S NESTED MODEL OF INFORMATION BEHAVIOUR RESEARCH AREAS (Source: Wilson, 1999)

Wilson noted that one constant complaint of commentators has been that researchers have not built upon prior research in such a way as to cumulate a body of theory and empirical findings that may serve as a starting point for further research. General information behaviour models rarely advance to the stage of specifying relationships between theoretical propositions. Behavioural models that provide a map of a theoretical area but provide no suggestion for causative factors make it hard for theories to be cumulated. For example, by using a flow-chart to represent information behaviour processes, it is possible to identify gaps in research (e.g., Wilson, 1981). However, flow-charts do not directly suggest a hypothesis to be tested because causative factors are often hidden.

From the perspective of the user, uncertainty is a factor that is always present. Wilson (1999) hypothesises that each stage of problem solving (identification, definition, resolution, solution) sees the successive resolution of more and more uncertainty, and where uncertainty fails to be resolved at one stage, it may result in a feedback loop to the previous stage for further resolution. Wilson believed that previous attempts at attempting to model information behaviour (e.g., Kuhlthau, 1991; Ellis, 1989) can answer the question 'how is the uncertainty resolved?' by viewing models as iterative steps that may occur in exploratory loops between each link in the problem resolution chain.

4.2.1 INTERFACE LEVEL

Beaulieu (2000) identified three principal aspects of interaction: interaction within and across tasks; the notion of interaction as task sharing; and interaction as discourse. The traditional model of the information retrieval task as a matching function between a query and a set of documents has slowly given way to the concept of extending the systems boundary to include the user as an essential component of the system (Robertson and Hancock-Beaulieu, 1992). It is suggested that HCI is primarily concerned with establishing a user/system dialogue at the user interface level and does not address the interactive nature of IR tasks. Beaulieu (2000) believed that HCI models such as Moran's (1981) are too generic and do not address the link between operational tasks (interface level) and the conceptual tasks need for interactive IR. Likewise, she suggests that descriptive models of information behaviours do not provide sufficient analysis depth for task interaction specification (at interface level).

During the late 1960s and early 1970s, most online systems provided a single level of support to all users (Savage-Knepshield and Belkin, 1999). At *The User Interface for Interactive Search of Bibliographic Databases Workshop*, Bennett (1971) challenged researchers to consider issues that needed to be addressed in order to provide more effective interfaces. He argued that different levels of user experience should be supported by enabling behavioural characteristics to be correlated with interactive display techniques. He also raised the issue of how feedback can be implemented in order to provide assistance during search formulation. Ellis (1989) noted that although a considerable amount of research has now addressed issues of interface design, there has been less research directed towards understanding the nature of interaction. The period from the mid-1970s to the mid 1980s was marked by a better understanding of a searcher's problem space (or information need) during the search process. The cognitive approach provided a new framework for exploring behaviour, for example, *see* Belkin (1980).

4.2.2 COGNITIVE LEVEL

Most past behavioural research in IR has focussed on expert users and intermediaries. In cognitive canonical models, a standard user model is assumed at the beginning of each session, rather than being built from scratch as the search progresses. Daniels (1986) argued that there is a limit to the usefulness of canonical modelling to systems with heterogeneous user communities or which do not have restricted fields of application. Many systems have too wide a field of application to be able to form a representation of the so-called 'ideal user'. Such systems should be able to adapt to individual users, rather than merely assigning each user to a certain predefined subset. However, without the ability to quantitatively measure changes in cognitive structures, it is impossible to provide a system that is capable of adapting effectively.

Cognitive models can be regarded as images that the components of a system, whether the components be people or machines, have of themselves, of each other, and of the world. These models enable individuals to make inferences and predictions, to understand phenomena, to decide what actions to take and to control its execution, and above all to experience events by proxy (Johnson-Laird, 1983). The important consideration here is to emphasise the fact that cognitive models are merely self-generated perceptions. These perceptions are based on a wide range of variables that include affective feelings as well as cognitive behavioural patterns. Daniels (1986) noted that analytical cognitive models attempt to model aspects of a user's cognitive behaviour in a more qualitative manor, and such aspects may include the user's knowledge of the underlying task, his goals, plans, beliefs, background, preferred learning and interaction style, system experience, etc.

Ingwersen's (1996) model of IR interaction is an ambitious attempt at an integrated analytical cognitive model. By identifying processes of cognition that may occur in all information processing elements (polyrepresentation), Ingwersen demonstrated that a comprehensive model of information behaviour must include the system that points to the information objects that may be of interest to the user. He thereby, incorporated the idea of information behaviour with IR system design.

4.2.3 USER MODELLING

Wilson (1999) argued that the whole of information behaviour should not be subsumed under the heading of the 'user's cognitive space'. Issues of how users arrive at the point of making a search, and how their cognitive structures are affected by the process of deciding how and when to move towards information searching may be lost. Spink (1997) argued that judgements made by a user must be based upon prior experience, gained in the overall activity of information seeking, and derived from behaviour that proves to be useful in settings other than the interactive IR system. The development of a general cognitive model that can encapsulate the process of problem solving (in the same way that Wilson views the process) cannot be achieved exclusively from a cognitive viewpoint.

Instead of attempting to model a user's information need cognitively, maybe it is more important to study a user's information behaviour. Implementing system features that enables behaviour to be supported, allows greater flexibility of interaction rather than trying to predict what features should be available to a user depending on their cognitive state. A user may well be unable to communicate his or her cognitive state to a system. A user's perceived mental model may not be an accurate reflection of an information need situation. Offering flexibility of search techniques may force a user to realise that a perceived mental model may be inaccurate.

According to Rich (1983), one approach to modelling is to let the users provide their own model, or explicitly create their own environments. Rich's (1979) GRUNDY system was an ambitious attempt that enabled users to inform the system of information requirements using a feedback mechanism. The GRUNDY system aimed to recommend novels to library users. The user has a predefined goal of finding a novel. The systems role is to construct a user model that can provide details of the user's likely interests. GRUNDY began a consultation by asking a user to enter a few self-descriptive terms that were then used to activate a number of stereotypes. These stereotypes enter values into the user model. The system then resolved any conflicts between different stereotypes by asking questions in an attempt to reduce uncertainty.

Depending on how confident GRUNDY was of the user model, the system either recommended a novel to the user or asked further questions. Interaction is a good means of ascertaining facts, resolving conflicts, assigning confidence ratings to the information goals, and updating / modifying the user model. The generation of stereotypes, used as a mechanism for interaction, is only suited to narrow domains of applicability where user interests can be simply inferred.

Boguraev (1985) believed that the analysis of interaction between people before online IR might be a good way of building a user model. Boguraev emphasised that the dynamic change in a user's problem state can only be achieved by continually maintaining and updating an individual's model, and that the only way of obtaining information about a user's perception of the problem is through natural dialogue. However, it must be recognised that language is not necessarily always a suitable method for communicating affective dimensions of information requirements. The communication of affective dimensions may well be transmitted using non-verbal actions. The level of uncertainty experienced by an individual, for example, cannot always be interpreted through the analysis of language alone. More invasive techniques may well be required.

Daniels (1986) believed that it is unlikely that the user's plans (i.e., preferred search strategy) can be successfully incorporated within the system's model of the user. Daniels claimed that most users do not have a plan of how the interaction should proceed[20]. Oddy's (1977) THOMAS program formed an image of the user's interests by allowing the user to browse. THOMAS took account of the non-specifiable nature of some individuals' information needs. THOMAS is a good example of an adaptive IR system. Bates (1990) believed that the objective of much research in IR is to design systems in which the user has only a reactive involvement with the search process (as opposed to an adaptive involvement). She argued that designers of IR systems are missing out on the opportunity to design systems that are in better harmony with the actual preferences of users.

[20] It can also be argued that users might have a plan but the nature of the information prevents the plan being executed (*see* Chapter 11).

The question of how much and what type of activity the user should be able to direct the system to do is posed by Bates. Information seeking strategies are composed of a number of complex interactions with information objects (e.g., monitoring, identifying, reformulating, etc.). Bates claimed that it is not possible to classify the types of information behaviour activity as belonging to a particular strategy. Sometimes a strategy can be composed of a single action, sometimes several. Bates believed that a strategy is an emergent phenomenon; each higher level of search activity is conceptually different from lower levels and is more extensive. By explaining the concept of information search strategies, Bates showed the importance of providing strategic support to the user rather than purely procedural capabilities (i.e., system features that require only a reactive response).

4.2.4 SITUATIONAL MODELS

Dervin (1983) suggested that information processing and use are situationally bound behaviours; they occur at specific points in time and space in order to fit the demands at those points in time and space. To predict communication-information behaviours, the predictor variables must be selected in a way congruent with the context within which these behaviours are selected (time and space) and consistent with the views of the person selecting the behaviours (individual). Dervin revealed that the use of the 'personality predictor approach', adopted by psychologists such as Carter (1974), consequently fail to predict information processing behaviour and use. Information seeking and use can be predicted more powerfully by knowing the kind of situations users are in rather than knowing their personality or demographic attributes.

IR can be seen as part of a broader process of information seeking. IR systems are a specific means among other sources and channels for obtaining information. It is obvious that the role of IR systems and the information they provide varies depending on the task and situation. Vakkari (1999) believed that studies on information seeking behaviour have not contributed to how changes in the problematic situation and information needs are reflected in patterning of search strategies and relevance assessments. Vakkari claimed that information activities (search strategies for obtaining information and determining relevance) are systematically connected to task complexity and the structure of the problem at hand (*see* Figure 4.2).

The more structured the problem, the more precisely the relevance of a document can be assessed in terms of content and meta-information. Kuhlthau (1993b) has shown that task performance can be differentiated into various phases depending on an individual's perception of their information need. She found that once a focus of the task had been identified, individuals employ a wider range of information seeking strategies. The structure of the problem faced influences the patterning of strategies. Task complexity can be understood in many ways. It has been associated with the pre determinability of, or uncertainty about, the task (Byström & Järvelin, 1995). This dimension is related to the following characteristics of a task: repetitivity, analysability, the number of alternative paths of task performance, and outcomes of novelty (Vakkari, 1999).

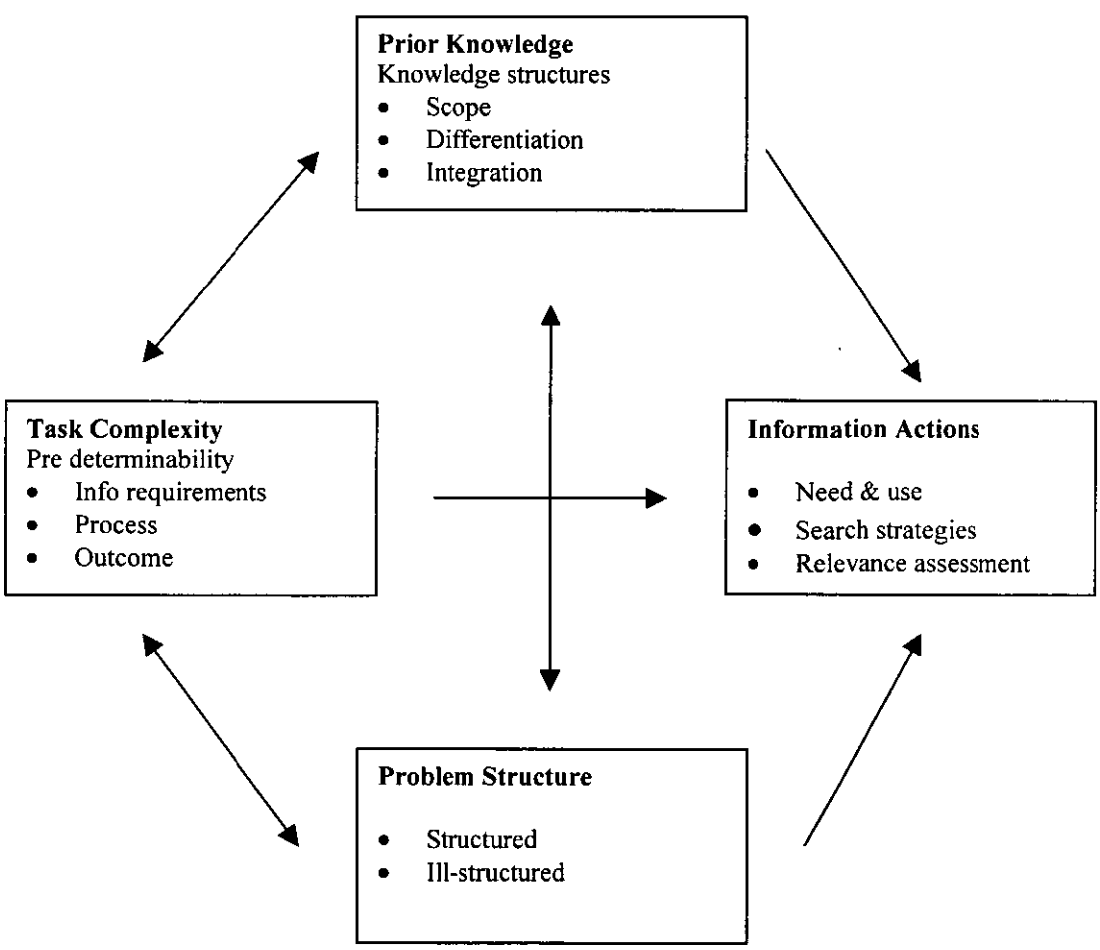


FIGURE 4.2: KNOWLEDGE, TASK COMPLEXITY, INFORMATION ACTIONS, AND PROBLEM STRUCTURE (Source: Vakkari, 1999)

Structured tasks allow a higher degree of pre-determinability. Tasks become more structured if an individual has prior knowledge. Vakkari believed that prior knowledge leads us to selectively attend to certain features during learning, thereby narrowing the problem space. Learning is most efficient when the structure to be learned is compatible with the structure that was expected according to prior knowledge (Heit, 1997). However, as Vakkari (1999) identified, one of the main difficulties in research has been how to describe changes in understanding, i.e., in the cognitive structures in a way that would connect the change process to changes in information actions.

Some critical problems associated with the nature of interaction have been identified by Ellis (1996). The development of a representation of an individual's requirements, especially when these requirements are dynamic, requires complex dialogue. If the dialogue between the user and the system is defective, then any information derived from that interaction would also be deficient and misleading.

Yoon and Nilan (1999) studied interpersonal information seeking interactions between a user and a human information source. They found that users specify their information need (uncertainty) largely in terms of what they know (certainty) during interaction. Taken together, certainty and uncertainty represent a gestalt of the user's perception of an information need. Yoon and Nilan claimed that the interaction between a user and a human source provides a model of information seeking interaction where meaning is shared more directly with little mediation or functional limitations inherent in current automated interactions with IR systems.

An information need has been defined as uncertainty (Dervin, 1983; Dervin & Nilan, 1986); an anomalous state of knowledge (Belkin, Oddy, and Brookes, republished 1997); a cognitive gap (Dervin, 1983; Dervin and Nilan, 1986); or a visceral need (Taylor, 1986) with the assumption that human cognition is a dynamic condition. **This thesis defines an information need as an emergent realisation, triggered by the uncertainty associated with a gap in knowledge. The gap in knowledge may be between an individual's current state of domain knowledge and the goal state, or between an individual's current state of procedural knowledge and the goal state. Both these gaps cause a cognitive load that influences a user's ability to interact with information objects (see Chapter 7).**

The concept of certainty, as defined by Yoon and Nilan (1999), primarily addresses what users already know regarding their need situation. They claimed that the concept of uncertainty is what a user needs to find; everything else is considered as certainty. This is a very narrow definition, as an individual can have multiple overlapping states of: uncertainty, anomalous states of knowledge, cognitive gaps, and visceral needs, associated with their: information need, ability to find information, ability to understand information, ability to compile a coherent solution statement, etc.

Yoon and Nilan's interpersonal study examines information seeking interaction without the perpetuating limitations in current IR systems. Establishing a better understanding of human behaviour during the exchange of meaning may enable what variables, beyond topical specifications, are needed to facilitate IR. However, it can be argued that assuming no perpetuating limitations exist during human-human interaction is dangerous. This thesis argues that humans themselves have perpetuating limitations due to their cognitive architecture. This concept is introduced in the following chapter (*see* Chapter 5). Although users define their information need in terms of what they know, the use of certainty as a measure of a user's perceptions at the time of interaction, requires an understanding of cognitive limitations.

4.3 OUTSTANDING ISSUES

- As knowledge states shift to more clearly focussed thoughts, a parallel shift occurs in feelings of increased confidence. Commonly experienced phases or stages provide an abstract view of the search process. They do not increase our understanding of how an individual progresses from one search stage to another. This thesis will avoid making the assumption that uncertainty is a suitable measure of an individual's ability to process information. The possibility that manifestations of uncertainty are symptomatic of limitations to cognitive capabilities needs to be considered.

- A number of models have been proposed that can be used to explicate user searching behaviour during the process of relevance feedback. Generally, models of information behaviour do not attempt to describe the same types of activities; each of them represents user interaction at different levels of abstraction. Cognitive models are merely self-generated perceptions. Without the ability to quantitatively measure changes in cognitive structures it is impossible to develop a cognitive model of a user that is capable of adapting to user needs.

- Instead of attempting to model a user's information need cognitively, maybe it is more important to study a user's information behaviour. Implementing system features that enable that behaviour to be reflected, allows greater flexibility of interaction rather than trying to predict what features should be available to a user depending on their cognitive state. A user may well be unable to communicate their cognitive state to a system. A user's perceived mental model may not be an accurate reflection of their information need situation. Offering flexibility of interaction may allow a user to realise that their perceived mental model may be inaccurate.

- To model behaviour, predictor variables must be selected in a way congruent with the context within time and space, and consistent with the views of the individual. Prior knowledge leads us to selectively attend to certain features during learning, thereby narrowing the problem space. Although users define their information need in terms of what they know, the use of certainty as a measure of a user's perceptions at the time of interaction, requires an understanding of cognitive limitations.

5.0 INTRODUCTION

Educational Psychologists have been researching human information processing since the 1950s. Very few studies conducted by information scientists have used these findings. This chapter introduces key concepts associated with information processing and problem solving. The relationship between learning and cognitive load is discussed. Many of the research findings have clear implications for the study of information behaviour. Attempts at artificially simulating / modelling the complex process of problem solving in heterogeneous domains have been unsuccessful. However, these attempts can provided useful insights into the nature of information behaviour.

5.1 COGNITIVE MODELS OF INFORMATION PROCESSING

As discussed previously (*see* Section 1.4), Miller (1956) introduced the idea that short-term memory (or working memory) has a limited capacity, holding 5-9 chunks of information. SOAR (Laird, Newell, and Rosenbloom, 1987) is an example of an attempt at a unified theory of cognition that builds on Miller's idea of chunking as the basis for the organisation of memory. Miller also introduced the influential TOTE (Test-Operate-Test-Exit) concept (Miller, Galanter, and Pribram, 1960). A TOTE unit measures problem solving effort. A goal is tested to see if it has been achieved and if not, an operation is performed to achieve the goal. This cycle is repeated until the goal is achieved or abandoned. TOTE replaces 'stimulus-response' as the basic unit of behaviour. The classic example for TOTE is a plan for hammering a nail. The Exit Test is whether the nail is flush with the surface. If the nail is not flush, then the hammer is tested to see if it is raised (otherwise it is raised) and the hammer is allowed to hit the nail. The concepts of chunking and TOTE are low level theories which cannot help analyse something as complex as information behaviour. However, Miller's theories have shown that the process of planning is the fundamental cognitive process that influences information processing.

ACT* is a general theory of cognition developed by Anderson (1983). This theory has been used as the basis of many attempts to design computer tutors (Anderson, Boyle, Farrell, and Reiser, 1987). Three types of memory structures are defined: declarative, procedural, and working.

Declarative memory takes the form of a semantic net that attempts to link elements of information. Procedural memory attempts to assign a set of conditions/rules to each element of information, generating a production. The ACT* principle works on the basis that all knowledge is generated from declarative information. Procedural knowledge is acquired by making inferences from existing declarative information. This approach is very simplistic. It is reliant on the use of logical constructs, which attempt to mimic the mental state of a user.

ACT* cannot represent the notion of contextual meaning, which renders it totally unsuitable for the examination of information seeking behaviour. However, ACT* does provide evidence that even when simplistic assumptions about the nature of cognition are applied, a variety of learning mechanisms emerge. ACT* supports three fundamental types of learning: generalisation, discrimination, and strengthening. Generalisation is where productions become broader in their range of application. Discrimination occurs when productions become narrow in their range of application. Strengthening arises when productions are repeatedly successfully applied. New productions are formed by the conjunction or disjunction of existing productions. Successfully applying the ACT* principles requires the following approach:

1. Identify the goal structure of the problem space
2. Provide instruction in the context of problem solving
3. Provide immediate feedback on errors
4. Minimise working memory load
5. Adjust the "grain size" of instruction with learning to account for the knowledge compilation process.
6. Enable the student to approach the target skill by successive approximation.

The concept of adjusting the "grain size" of instruction to facilitate learning by reducing cognitive load is intriguing. Adjusting "grain size" is achieved by a feedback loop. Within an IR system, this type of feedback cannot be easily achieved. Assigning an element of information with an appropriate feedback tag would be difficult. Users would have to identify to what extent an element has caused generalisation, discrimination, or strengthing. Even if this was possible, there is no evidence that retrieving information, which facilitates the minimisation of cognitive load would assist the information seeking process. It is likely that retrieving information that is compatible with existing knowledge, especially in the early stages of the search process, would constrain the development of an information need.

Cognitive flexibility theory is a constructivist theory that was specifically formulated to investigate the use of interactive technology, e.g., hypertext. Spiro and Jehng (1990) were the initial pioneers of the theory. They suggested that cognitive flexibility is the ability to spontaneously restructure knowledge in response to situational demands. When using an IR system, a user needs to be able to spontaneously restructure knowledge as an information need develops. The dynamic nature of an information need means that if knowledge can be represented along multiple rather than single conceptual dimensions, then the assembly of new schemata is aided. Cognitive flexibility theory is largely concerned with the transfer of knowledge and skills beyond their initial learning situation. The theory asserts that effective learning is context-dependent and learners must be given the opportunity to develop their own representations of information in order to properly learn/understand. Experimentation performed by Jonassen, Ambruso, and Olesen (1992) involved presenting a number of different clinical cases to medical students. Participants were required to make a diagnosis using various sources of information provided by a hypertext system. It was found that if information sources provided multiple representations of content, avoided oversimplification, and were highly interconnected rather than compartmentalised, then effective learning would occur. Unlike ACT*, the principles of cognitive flexibility theory can be easily applied to the analysis of information seeking behaviour (*see* Chapter 11).

B 011309

– will be back

in about 30mins.

Matt

17·21 / 9/8/11

Festinger's (1957) cognitive dissonance theory stated that there is a tendency for individuals to seek consistency among their cognitions (i.e., beliefs, opinions). If an inconsistency or dissonance occurs, attitude has to change to accommodate behaviour, or behaviour has to change to accommodate attitudes. If dissonance occurs during IR interaction, then it is unlikely that information seeking behaviour will change unless the strength of the dissonance increases a user's level of uncertainty to a level that affects the cognitive load imposed on working memory. The greatest dissonance occurs when two alternative approaches to the information problem are equally feasible. Attitude change is more likely to be in the direction of information that is less threatening, since this results in lower dissonance and minimises cognitive load. Cognitive dissonance theory is therefore contradictory to most behavioural theories, which predict greater attitude change when novel information is encountered. Identifying if dissonance has affected information seeking behaviour is difficult since the information elements generating the dissonance cannot be easily identified. Information elements exist in declarative and procedural memory and are not fully integrated within working memory. This crucial limitation to information processing should be considered when analysing information behaviour (*See* Chapter 6).

Bartlett (1932, 1958) was the first to suggest that a schema provides a mental framework for understanding and remembering information. He performed experimentation on memory and found that participants recalled details of stories that were not actually incorporated within the stories. Chi, Glaser, and Farr (1988), amongst many others, have discovered that the nature of expertise is largely due to the possession of schemata that guide perception and problem solving. The schema construct is crucial to cognitive load theory and many other theories of cogitation. This thesis argues that schemata are the most suitable psychological construct for examining information seeking behaviour because they allow for the emergence of meaning (*See* Chapter 7).

Bruner (1973) argued that learning is an active process in which new ideas are constructed based upon current and past knowledge. Cognitive structures provide meaning and allow an individual to 'go beyond the information given'. An individual selects and transforms information, constructs hypothesis, and makes decisions, relying on a cognitive structure to do so.

The ultimate goal of an IR system should be to translate information into a configuration appropriate to an individual's current state of understanding. Bruner's constructivist theory provides a framework for the development of instruction within learning environments. Bruner (1986, 1990) believed that:

- Instruction must be concerned with the experiences and contexts that make a student willing and able to learn.
- Instruction must be structured so that it can be easily grasped by the student.
- Instruction should be designed to facilitate extrapolation (i.e., allowing a student to go beyond the information given).

Bruner's ideas clearly lend themselves to the IR process, and should be incorporated into models of information behaviour (*see* Chapter 12):

- Information behaviour must be concerned with the experiences and contexts that enable an individual to successfully retrieve information.
- IR should ideally be a structured process so that it can facilitate an individuals understanding.
- The analysis of information behaviour should only take place at the epistemic or behavioural level of information processing.

Metacognition is the active monitoring and regulation of one's own cognitive processes. Flavell (1976) suggested that "I am engaging in metacognition if I notice that I am having more trouble learning A than learning B; or if it strikes me that I should double check C before accepting it as fact." (p. 232). Metacognition is relevant to work on cognitive styles in so far as the individual has some awareness of their thinking process. Cognitive styles refer to the preferred way an individual processes information. Styles are considered to be bipolar dimensions, whereas abilities are unipolar. Possessing more of an ability is considered beneficial while having a particular cognitive style denotes a tendency to behave in a certain manner.

Theories of field dependent and field independent cognitive styles have a number of implications for information seeking behaviour and will be discussed further in Chapter 12. Witkin and Goodenough (1981) influential study identified a number of other cognitive styles that also have an impact on learning:

- *Scanning* – Differences in the extent and intensity of attention.
- *Levelling versus Sharpening* – Differences in remembering distinct memories versus the tendency to merge similar events.
- *Reflection Versus Impulsivity* – Speed and adequacy with which alternative hypotheses are formed and responses made.
- *Conceptual differentiation* – Tendency to categorise perceived similarities among stimuli in terms of separate concepts.

Although it is possible to observe field dependent and field independent behaviour during information seeking, the above cognitive styles cannot be observed directly. Categorising participants as possessing a particular cognitive style is often difficult. For example, Kolb (1984) proposed a theory of experiential learning that involves four stages: concrete experiences (CE), abstract conceptualisation (AC), active experimentation (AE), and reflective observation (RO). The CE/AC and AE/RO dimensions are polar opposites. Depending on these two dimensions, four types of learner exist: divergers, assimilators, convergers, and accommodators. Divergers prefer CE and RO. Assimilators prefer AC and AE. Convergers prefer AC and RO. Accommodators prefer CE and AE. Attempts to explain behaviour by qualitatively categorising individuals prevents an understanding of interrelations between behavioural factors and may lead to inaccurate generalisations.

Many previous studies have not attempted to model interrelations. However by applying a quantitative approach, Heinström (2002) estimated the influence of a number of personality and learning style factors on information seeking behaviour. Three different patterns of information behaviour were found: fast surfers, broad scanners, and deep divers. Psychological mechanisms other than cognitive differences are connected to information seeking and need to be represented within a model of behaviour. Höglund (2003) suggested that Heinström's study showed that a combination of both quantitative and qualitative approaches is needed to reveal the complex relationships between factors.

5.2 COMPUTER SIMULATION OF PROBLEM SOLVING

Theories of human cognitive processes can be attempted at the level of neural processes, elementary information processes (e.g., retrieval from memory), or at the level of higher mental processes (e.g., problem solving). This thesis does not deal with neural processes or elementary processes. The proposed cognitive load theory for relevance feedback does not attempt to quantify information processes by using algorithms or computational units. Cognitive load theory attempts to model the functions of higher mental processes that can be associated with the limitations of empirically verifiable cognitive attributes (i.e., working memory) (*See* Chapter 6). It is interesting, however, to examine attempts at developing theories of problem solving that can be simulated using computers. Such attempts reveal some valuable insights into the complexity of human information behaviour.

The period from 1956-1972 saw the emergence of a well developed theory of problem solving, and theories of the performance of two types of induction tasks: concept attainment and sequence extrapolation (Simon, 1981). Even in relatively simple problem domains, a number of alternative problem solving strategies are possible (Greeno, 1976). Some of these strategies depend on perceptual clues. Barry (1994) discussed the potential of using clues to identify characteristics of documents that could be used by an IR system to retrieve documents (i.e., going beyond topicality). However, the interpretation of perceptual clues are dependent on structures of goals held in memory and expertise (Chase and Simon, 1973). Importantly, identifying the characteristics of perceptual clues, where the discovery of novelty is an essential element, cannot be achieved from post-hoc generalisations (*see* Chapter 12).

In a problem space, various individuals may use different problem solving strategies. Simon and Reed (1976), amongst others, found that a wide range of behaviour could be accounted for by applying a means-ends strategies with consideration for iteration, but little need for planning. However much research on problem solving has been performed using tasks that do not require specific domain knowledge. A semantically rich domain, such as chess, requires specific knowledge and problem solving abilities (Chase and Simon, 1973). Rieger (1976) questioned how knowledge could be organised and accessed to enable expert behaviour. By making empirically testable assumptions about the nature and organisation of short term memory and long term memory, it is possible to build production systems to simulate behaviour.

Such systems enable a basic rule of interpretation. Whenever conditions of a production (instruction / rule) are satisfied, the action of that production will be executed. Additional rules determine which production will fire when the conditions of several are satisfied simultaneously. Newell (1973) found that production systems are much more flexible at controlling a sequence of thought than traditional hierarchical computer programming languages. However, modelling a user's information behaviour is not possible using production systems. Before an individual can attempt to solve a problem, an understanding of the problem must be present (Greeno, 1977).

Information processing models of the problem-solving process generally view an input into the system as a formalised representation of a fully defined problem space (Hayes and Simon, 1974). These models also assume that an initial understanding process has run to completion. These type of assumptions are no longer being made by the majority of information behaviour studies. Treating an individual's IR query as an accurate representation of a fully defined problem space is nonsensical. The process of information seeking involves an individual's problem space becoming better defined during interaction with information objects.

A theory of long term memory must account for both the organisation of information stored and the routes by means of which information is accessed (Simon, 1981). The schemata of semantic memory may have either a topical or episodic character (Watkins and Tulving, 1975). Topical schemata could represent objects or concepts while episodic schemata could represent a system of causally related events. A node link in memory may have aspects of both topical and episodic schemata. Assuming that the memory structures of an individual can be represented in this way, it seems feasible to undertake the task of specifying processes for creating those memory structures. A problem solving program has to be motivated to solve problems. Likewise, an individual has to be motivated to use an IR system.

Colby's (1971) PERRY simulation introduced numbers to represent affective emotions. A two-way linkage between the levels of affect, represented as numerical quantities, and the symbolic cognitive content of long term memory, enabled paranoid responses to be triggered. If reference is made to sensitive topics, then affect levels are increased and may result in a verbal attack on the interlocutor (i.e., representing anger), or by withdrawal (i.e., representing fear). If such a model is used to represent the concept of uncertainty within information behaviour, it may be possible to simulate the triggering of a learning process. Interacting with information objects that are deemed by a system to be above a certain threshold of uncertainty (e.g., if they are poorly understood) could result in the transfiguration of existing topical and episodic schemata in an attempt to lower the uncertainty. Unfortunately, because we currently have little knowledge of how uncertainty affects information behaviour, it is impossible to speculate how uncertainty triggers transfigurations in schemata. There has been little research effort aimed at encompassing motivational and emotional processes in information processing models. Hence, computer simulation continues to be unsuitable for the modelling of information behaviour.

5.3 OUTSTANDING ISSUES

- The process of planning is the fundamental cognitive process that influences information processing. The dynamic nature of an information need means that if knowledge can be represented along multiple rather than single conceptual dimensions then the assembly of new knowledge schemata is aided. If dissonance occurs during IR interaction, then it is unlikely that information seeking behaviour will change unless the strength of the dissonance increases a user's level of uncertainty to a level that affects the cognitive load imposed on working memory (*see* Chapter 7).

- Information elements exist in declarative and procedural memory and are not fully integrated within working memory. This crucial limitation to the processing of information should be considered when analysing information behaviour (*see* Chapter 6). The interpretation of perceptual clues are dependent on structures of goals held in memory and expertise. Importantly, identifying the characteristics of perceptual clues, where the discovery of novelty is an essential element, cannot be achieved from post-hoc generalisations (*see* Chapter 11).

- We currently have little knowledge of how uncertainty or different cognitive styles affect information behaviour. It is currently impossible to speculate how uncertainty or adopting a particular cognitive style triggers transfigurations in schemata (*see* Chapter 11).

6.0 INTRODUCTION

Throughout this thesis, it has been speculated that the cognitive load imposed on an individual affects feedback judgements. Within the field of educational psychology, Cognitive Load Theory was developed to explain how cognitive load acts as a barrier to problem-solving performance. Information seeking can also be viewed as a problem-solving process (Kuhlthau, Spink, and Cool, 1992). However unlike the majority of problems studied by educational psychologists, information seeking does not have a pre-defined goal state. As previously discussed, for relevance feedback purposes, assessing and capturing relevance is highly problematic (*see* Section 2.3.1 & Chapter 3). The purpose of an IR system is to provide documents that are relevant[21] to an individual's information need. Examining cognitive load requires an understanding of the situational and behavioural context in which judgements occur.

This chapter reviews Cognitive Load Theory research in an attempt to establish how the theory can be extended to enable the evaluation of relevance feedback. Since humans have a limited capacity of assimilating new information, they purposefully construct meaning by selectively attending to that which connects with what they already know. The ability to assimilate information is burdened by cognitive load. Schamber, Eisenberg, and Nilan (1990) concluded that relevance is dependent on both *intrinsic* (cognitive/behavioural) and *extraneous* (situational) factors. Although Cognitive Load Theory cannot be used to measure relevance, it could enable a greater understanding of how relevancy judgements are influenced by quantifiable *intrinsic* and *extraneous* factors. Cognitive Load Theory is able to represent both *intrinsic* and *extraneous* elements during a *communication process*.

---

[21] Saracevic (1997) suggested that relevance could be considered as a measure of the effectiveness of a contact between a source and destination in a *communication process*.

The term 'cognitive load' has been used loosely within IR research, mostly in reference to Human Computer Interaction (HCI) issues. Although the term has been utilized frequently, the only IR study that attempted to define the concept was performed by Hu, Ma, and Chau (1999). They examined the effectiveness of designs (graphical or list-based) that best supported the communication of an object's relevance. Cognitive load was used as a measure of information processing effort a user must expend to take notice of the visual stimuli contained in an interface and comprehend its significance. It was assumed that users would prefer an interface design that requires a relatively low cognitive load and at the same time, can result in high user satisfaction. A self-reporting method was used to obtain individual user assessments of the cognitive load associated with a particular interface. The focus of this study was interface design, so the use of the term 'cognitive load' was valid from a HCI viewpoint. However, as will be demonstrated, the concept of cognitive load during IR can be extended far beyond interface design.

6.1 COGNITIVE LOAD THEORY AND PROBLEM SOLVING

This thesis is concerned with the development of a Cognitive Load Theory that is appropriate for the study of information behaviour. The proposed theory is considerably different from Sweller's (1988) original manifestation. However, because the ideology of the original theory has been preserved, the proposed theory should be considered as an extension rather than a revision. Sweller's original conception suggested that problem-solving ability[22] is enhanced under conditions that are aligned with human cognitive architecture. Although the structure of human cognitive architecture is not known, some traits are discernible through the results of experimental research in psychology.

[22] The process of problem solving and learning are substantially different, especially within the context of information seeking, but at this introductory stage no delineation is required.

Miller (1956) found that short-term memory (working memory) is limited in the number of information elements that it can hold simultaneously. Sweller, amongst many others, constructed theories based upon known (experimentally verifiable) limitations of cognitive architecture.

An individual's knowledge base is composed of complex cognitive structures known as schemata[23] (Chi, Glaser, and Farr, 1988; Sweller, 1988). Long-term memory contains schemata that have been acquired over a lifetime of learning. Schemata allow us to assimilate information through recognition. Problem solving occurs when existing schemata are successfully modified. Rather than grouping rote-learned facts, schemata permit multiple elements of information to be treated as a single element. The difference between an expert and a novice is that a novice hasn't acquired the schemata of an expert.

Chase and Simon (1973) found that both novices and experts remembered the positions of chess pieces in chunks, and the number of chunks did not differ appreciably. However, significant differences occurred in chunk size with experts possessing larger (more sophisticated) chunks than those of novices. This finding has been widely replicated in other domains such as learning to use computer programs (Cooper and Sweller, 1987). This type of research confirms that schemata for problem state configurations vary in complexity. Novices have less well-developed schemata.

The process of problem solving requires a modification of the semantic structure of long term memory. This is easily demonstrated by comparing novices with experts. Larkin et al (1980) discovered that strategies employed by expert and novice problem solvers differed when attempting to solve physics problems. Novices worked backward from the goal and set sub goals (means ends analysis). Experts, in contrast, were able to work forward by choosing appropriate actions leading to the goal as they recognised each problem state from previous experience and know which moves are appropriate.

[23] Sometimes referred to as schemas.

Well-developed schemata enables a problem solver to know that certain problems can be grouped, at least in part, by their similarity and the similarity of the moves that can be made from those states (Sweller, 1988). Novices are often error-prone and slow, while experts are methodical and fast. The change in performance occurs because as a learner (problem solver) becomes increasingly familiar with the material, and the cognitive complexity associated with the material is minimised (through the acquisition of improved schemata) so that it can be handled more efficiently by working memory.

The same principles (i.e., the possession of better schemata) leading to expert-novice distinctions in problem solving strategies, and memory of problem-state configurations, may also be instrumental in identifying differences in information behaviour (*see* Chapter 11).

Cognitive Load Theory is based on the idea that there is a memory architecture that consists of a limited working memory that interacts with an unlimited long-term memory (Baddeley, 1992). The limitations of working memory can be circumvented by coding multiple elements of information as one element in cognitive schemata. Working memory is used for all conscious activities and it is the only memory that can be monitored. As noted earlier, it is limited to about seven items (Miller, 1956). Everything else is concealed until brought into working memory. Working memory is used to organise, contrast, compare, and store information, which means that individuals can only process two or three items simultaneously (Baddeley, 1992). What constitutes as being an element (or item) of information depends on the schemata held by a person. A single element consisting of a single schema for an expert may be several elements consisting of several sub-schemata for a novice.

Long term memory is, in contrast, what individuals use to interpret and give meaning to experiences. It is a permanent repository of knowledge and skills and includes everything an individual has learnt. Schemata categorise information elements according to how they will be used (Chi, Glaser, & Rees, 1982). A key point to note is that although working memory can process only a limited number of elements at a time, the size, complexity, and sophistication of elements is not limited.

Cognitive Load Theory is concerned with the limitations of working memory capacity and the measures that can be taken to promote problem solving. Both causal and assessment factors affect cognitive load (Paas and Merriënboer, 1993). Causal factors are the characteristics of an individual (e.g., cognitive abilities, and domain knowledge), while assessment factors are characteristics of the mental effort needed and the mental load imposed. Mental effort refers to the cognitive capacity actually allocated to the task. Mental load is the portion of cognitive load that is imposed exclusively by the task and environmental demands. The individual's performance is a reflection of mental load, mental effort, and causal factors (*see* Figure 6.1, adapted from Paas and Merriënboer, 1993).

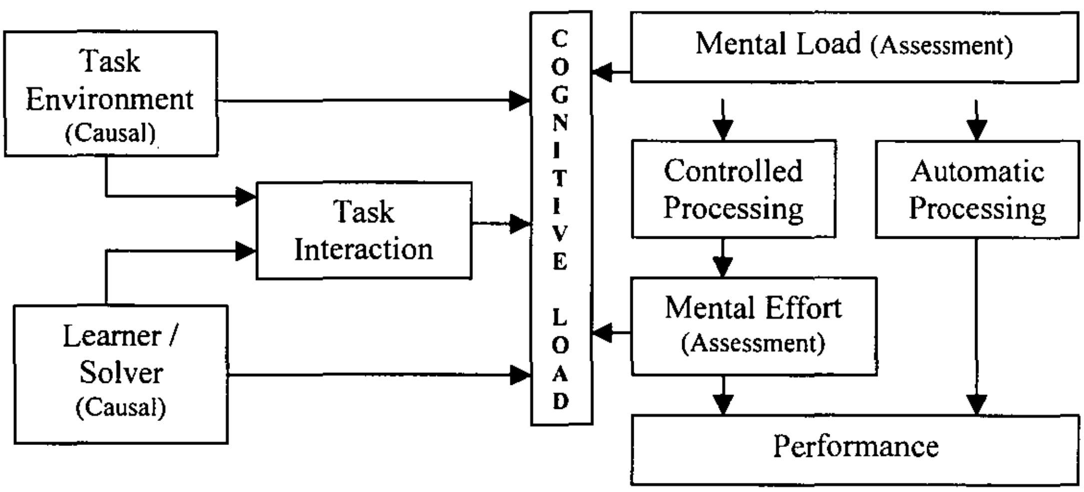


FIGURE 6.1: FACTORS DETERMINING THE LEVEL OF COGNITIVE LOAD

Cognitive Load Theory requires the representation of both the intrinsic (assessment) and extraneous (causal) aspects involved during problem solving. Within the field of instructional design, intrinsic load is induced by task difficulty, while extraneous load is induced by the nature/suitability of instruction (Brünken, Plass, and Leutner, 2002). In other words, intrinsic load is induced by an individual's expertise and the complexity associated with the task, while extraneous load is induced by learner activity (Gerjets and Scheiter, 2002).

If intrinsic load is high, and extraneous load is also high, then problem solving may fail to occur (*see* Figure 6.2):

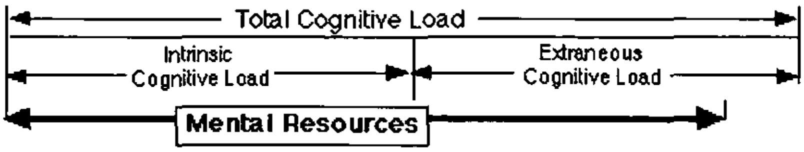


FIGURE 6.2: PROBLEM SOLVING FAILURE

When intrinsic load is low, then sufficient mental resources may remain to enable problem solving from any type of instruction (or information source), even if a high level of extraneous load is induced by the nature/suitability of instruction (or information) (*see* Figure 6.3):

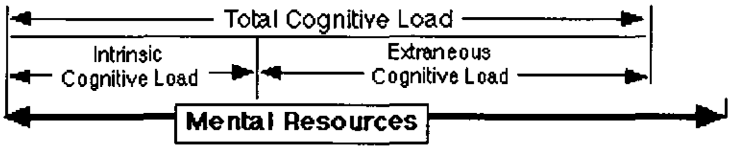


FIGURE 6.3: SUFFICIENT MENTAL RESOURCES

Modifying the instruction presentation (or information complexity) to a lower level of extraneous load will facilitate problem solving only if the total load falls to a level within the bounds of mental resources (*see* Figure 6.4).

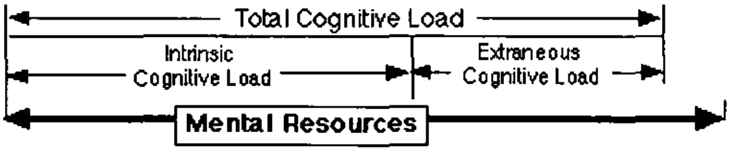


FIGURE 6.4: MODIFICATION OF EXTRANEOUS LOAD

Traditionally, Cognitive Load Theory research has studied methods that facilitated a decrease in extraneous cognitive load, allowing more effective problem solving. Instructional / educational material across a wide range of scientific disciplines has been modified in ways that allow the minimisation of extraneous load (Ayres, 1993; Paas, 1992; Van Merrienboer and Krammer, 1987; Cooper and Sweller, 1987; Sweller, Chandler, Tierney and Cooper, 1990). These extensive studies have led to a number of conclusions that are possibly universally applicable to all domains:

1. Change problem solving methods to avoid means-ends approaches that impose a heavy working memory load, by using goal-free problems.
2. Eliminate the working memory load associated with having to mentally integrate several sources of information by physically integrating those sources of information.
3. Eliminate the working memory load associated with unnecessarily processing repetitive information by reducing redundancy.

Cooper (1998) summarised the findings of research using Cognitive Load Theory and found that many conclusions were often counter-intuitive to standard educational practice (*see* Table 6.1).

Standard practice	**Cognitive load generated effect**
1. Use conventional problems which specify the goal so that students 'know what they have to find'.	The goal free effect Use goal free problems.
2. Students need to solve many problems to learn because 'practice makes perfect'.	The worked example effect Students learn by studying worked examples. Problem solving is used to test if learning has been effective.
3. Instructional materials which require both textual and graphical sources of instruction should be presented in a 'neat and tidy' fashion where the text and graphics are located separately.	The split attention effect Instructional materials which require both textual and graphical sources of instruction should integrate the text into the graphic in such a way that the relationships between textual components and graphical components are clearly indicated.
4. The same information should be presented in several different ways at the same time.	The redundancy effect Simultaneous presentations of similar (redundant) content must be avoided.
5. Similar to-be-learned information should be presented using an identical media format to ensure consistency in the instructional presentation.	The modality effect Mix media, so that some to-be-learned information is presented visually, while the remainder is presented auditorily.

TABLE 6.1: COGNITIVE LOAD GENERATED EFFECTS (Source: Cooper, 1998)

6.2 EXPANDING THE SCOPE OF COGNITIVE LOAD THEORY

Sweller's (1988) theory was proposed from the viewpoint of instructional design. Sweller's intention for the scope/applicability of his theory was in the domain of complex and technically challenging material. He was primarily concerned with the reasons that people have difficulty in learning from material of this nature. From this instructional perspective, information contained within instructional material must be processed by working memory. For schema acquisition to occur, instruction should be designed to minimise working memory load. Sweller's Cognitive Load Theory is concerned with the techniques for reducing working memory load in order to facilitate the changes in long term memory. Other researchers have attempted to extend Sweller's theory, especially within more language-based discursive areas (e.g., Brünken, Plass, and Leutner, 2002; Gerjets, and Scheiter, 2002). **The proposed extension to this theory, within this thesis, uses Cognitive Load Theory as a means of examining behaviour, rather than being concerned with suggesting techniques for minimising extraneous cognitive load.**

The primary reason why an individual undertakes information seeking is to facilitate understanding. Pollock, Chandler, and Sweller (2002) admitted that Cognitive Load Theory, as currently formulated, does not address the processing of highly complex information adequately, especially in language-based discursive areas such as IR. Methods of instruction, which are intended to facilitate complex understanding rather than problem solving, tend to incorporate all the information elements required. Pollock et al found that these types of instruction can overwhelm a learner's limited working memory and hinder learning. Artificially reducing the element interactivity of complex information allows elements to be processed serially. Results from Pollock et al's study indicated that for certain groups of learners, information is better understood through isolating interacting elements. This finding is of considerable importance to the study of information behaviour. It may help explain one of the reasons why individuals undertake different information seeking (or browsing) styles (*see* Chapters 7 & 12).

The issues associated with learning intellectually difficult material cannot be addressed in the same way as problem solving. Automation allows schemata to be processed automatically rather than consciously in working memory thus reducing cognitive load (Kotovsky, Hayes, and Simon, 1985). Intrinsic cognitive load of information is determined by the extent to which various elements interact. Some information can be understood independently of other elements. For example, the Spanish word for *bird* can be assimilated and learned independently of the word for *cat*. Therefore, the information is low in element interactivity. Only a limited number of elements need to be processed in working memory at any given time. High-element interactivity material consists of elements that cannot be understood in isolation because they interact. It is possible to learn vocabulary items individually but it is not possible to learn grammatical syntax without considering several vocabulary items and their relations.

Pollock, Chandler, and Sweller (2002) argued that the intrinsic cognitive load of instructional material is fixed if immediate understanding is the goal of the task. However, intrinsic cognitive load can be altered if a learning situation is created where understanding is not an objective. If a learner possesses under-developed schemata, and information consists of a large number of elements that are high in element interactivity, it will not be possible to process all elements simultaneously. For novices, in initial stages of schema construction, studying isolated elements followed by interacting elements would be more beneficial to learning than repeated exposure to the interacting elements of material. This prediction can be supported by research into learning hierarchies that suggested that lower level sills are prerequisites to higher order skills (e.g., Gagne, 1970).

Experimentation was conducted by Pollock, Chandler, and Sweller within the field of electronics, using first year electrical and mechanical university students. A range of practical tasks were set and students were randomly assigned to two groups. One group were presented with instructional material that enabled elements to be studied in isolation and then presented in a way that required an understanding of how these elements interacted. The other group were exposed to instructions that were devised to prevent elements being understood independently.

Results showed that when the interactivity between elements is reduced, discrete information units are created. Serial rather than simultaneous processing of information elements can then be undertaken, substantially reducing working memory load. The conclusion that understanding and learning of complex information can be facilitated by presenting the information that initially precludes full understanding is likely to be seen as controversial. Further research is needed within a wide range of subject domains.

Cognitive Load Theory applied to the domain of problem solving provides a pragmatic insight into how information can be presented in a way that circumvents cognitive limitations. When Cognitive Load Theory is applied to the domain of learning, fewer pragmatic insights are likely as the interpretation of meaning affects understanding in ways that cannot be modelled by Cognitive Load Theory (*see* Chapter 12). However, Cognitive Load Theory can be a useful framework for the exploration of behavioural processes during learning.

6.3 COGNITIVE LOAD THEORY AND LEARNING

Sweller (1988) believed that domain knowledge in the form of schemata is the primary factor distinguishing experts from novices in problem solving. Findings from research into expert-novice behaviour can help to explain why some forms of problem solving behaviour interfere with learning. He suggested that a major reason for the ineffectiveness of problem solving as a learning device is that the cognitive processes required by the two activities overlap insufficiently, and that conventional problem solving requires a relatively large amount of cognitive processing capacity which is consequently unavailable for schemata acquisition. Sweller showed that some forms of problem solving such as means end analysis interfere with learning. He believed that solving a problem and acquiring schemata might require largely unrelated cognitive processes.

In order to acquire a schema, a problem solver must recognise that a problem state belongs to a particular category of problem states that require particular moves. It is logical to assume that the attention to problem states previously arrived at, and the moves associated with those states, are important components of schema acquisition. Simultaneously considering the current problem state, the goal state, the relation between the current problem state and the goal state, the relations between problem solving operators and a goal stack (sub goals) impose a heavy cognitive load.

The safe assumption is that the satisfaction of an information need requires at least some learning needs to be met. A Cognitive Load Theory suitable for the study of information behaviour needs to be able to identify the types of search strategy or procedures that impede the satisfaction of an information need. Any potential method must be capable of simultaneously accounting for problem difficulty, subject knowledge, and strategy used (Sweller, 1988). There is much evidence that information seeking is directed by affective emotions, such as uncertainty (Ingwersen, 1992). These emotions may directly affect an individual's processing capacity by instigating particular forms of information seeking behaviour that, in turn, affect cognitive capacity (*see* Chapter 7).

For example, an individual who has a low level of domain knowledge is unlikely to be able to identify potentially useful information as well as a domain expert. During information seeking, the overall aim is to find information that enables the reduction of affective anxieties such as uncertainty. However, performing actions that result in an immediate reduction in affective anxieties may have a negative affect on satisfying an individual's information need (i.e., purposefully avoiding conflicting or complex information may hinder learning). Exposure to threatening, contradictory, or complex information is likely to increase levels of anxiety. Kuhlthau (1993a) suggested that learning to tolerate uncertainty as being a normal experience during information seeking is important. A domain expert is likely to use a schema-driven approach to information seeking. Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves, reducing cognitive load.

Exploration of the problem space to determine possible moves is a useful tactic during problem solving (Sweller, 1988). However, exploration of a problem space during information seeking is essential as the goal state is always changing due to an information need being a dynamic phenomenon. A Cognitive Load Theory for information behaviour needs to take into account this fundamental difference between problem solving and learning.

Extraneous cognitive load is the effort required to process poorly designed instruction, whereas *germane cognitive load* is the effort imposed by instruction that contributes to the construction of schemata (Kirschner, 2002). Germane cognitive load is required for the construction and storage of schemata into long-term memory, i.e., learning (Sweller, van Merriënboer, and Paas, 1998). The basic assumption is that an instructional design that results in unused working memory capacity because of a low intrinsic cognitive load and/or a low extraneous cognitive load may be further improved by encouraging learners to engage in conscious processing that is directly relevant to schema induction (Sweller, 1999).

Increasing germane cognitive load is a rather vague process. Researchers have currently found no empirical evidence or technique that allows discrimination between extraneous and germane cognitive load. (Van Merriënboer, Kirschner and Kester, 2002). Brünken, Plass and Leutner (2002) claimed that the aim of instructional design is to optimise load by decreasing extraneous cognitive load and increasing germane cognitive load. While the concept of germane cognitive load is intriguing, it is currently a tentative proposition that will not be integrated within the Cognitive Load Theory proposed by this thesis.

6.4 COGNITIVE LOAD THEORY AND BEHAVIOUR

Kirschner (2002) reviewed a wide range of papers that use Cognitive Load Theory. He believed that the theory offers promising perspectives on how instructional materials for acquiring cognitive skills and competencies should be designed. However, Kirschner warned that intrinsic load is not really a fixed quality because it is affected by causal factors.

These causal factors identified by Kirschner are the same factors studied by many researchers who are interested in explaining information behaviour. Dervin (1983) suggested information behaviour could be viewed as a process of sense-making. Sense-making involves many causal factors. An individual is actively involved in finding meaning that fits in with what they already know within a personal framework of reference. Kuhlthau (1991) argued that a model representing a user's sense-making should incorporate: physical, actual actions taken; affective, feelings experienced; and cognitive, thoughts concerning both process and content. Since people have a limited capacity of assimilating new information, they purposefully construct meaning by selectively attending to that which connects with what they already know. The ability to assimilate information is burdened by cognitive load.

Kirchner (2002) is concerned that the role of prior knowledge (domain knowledge) with respect to cognitive load should not be overlooked. Learners with deficient domain knowledge may adopt a more meticulous approach that enables the development of a *metacognitive control* facilitating the learning process. Kelly's (1963) personal construct theory is an attempt to explain the affective experiences of individuals. Kelly suggested that when new information is assimilated, confusion increases as inconsistencies and incompatibilities are found. If the information becomes too threatening, it is discarded. Alternatively, forming a tentative hypothesis may be possible in order to test whether information can be assimilated. Kelly's theory can be seen as a type of metacognitive control.

Metacognition is the active monitoring and regulation of one's own cognitive processes. A number of studies have shown that interest in subject matter has specific effects on the metacognitive control (Stark et al, 2002). Individuals with high levels of motivation make greater use of deep learning strategies. Deep learners are able to organise and elaborate information and monitor their understanding at the same time (Hidi, 1990). Stark et al (2002) found that successful learners more frequently verbalised that they failed to understand example information than the less successful, who more frequently signalled understanding. The meta-cognitively optimistic learners, who were less successful at learning, were victims of 'illusions of understanding'.

Stark et al believed that some individuals with a high tolerance of ambiguity tend to approach unknown, inconsistent, and complex situations as a positive challenge. Those with a low tolerance are inclined to interpret ambiguous situations as a threat. Different coping strategies may emerge as a result of an individual's tolerance. Sorrentino et al (1998) found that individuals with a high tolerance tend to process information deeply, while those with low tolerance adopt a mechanical or superficial information processing behaviour. Stark et al (2002) found a high tolerance of ambiguity concurred with high *mental effort* and resulted in effective, meta-cognitively accentuated learning. Mental effort is the subjective component of cognitive load and the central concept in Cognitive Load Theory. Differences in the effectiveness of learning methods can be explained by the mental effort of learners (Paas, 1992).

Stark et al did not employ the concept of mental effort as an independent variable. Mental effort is conceptualised as a subjective correlate. It can have effects on behaviour, but also itself can be influenced by the task. Stark et al make the assumption that the behaviour of learners is mirrored by their mental effort; deep learners are probably putting in more effort. Their focus was to investigate to what extent learner characteristics are responsible for differences in mental effort and learning behaviour, rather than attempting to measure mental effort by charging it to a particular account such as intrinsic cognitive load or extraneous cognitive load.

All the causal factors mentioned above that affect learning from instructional material also affect information behaviour. Factors commonly investigated include prior knowledge (domain knowledge), interest in the subject matter, and the tolerance of ambiguity (Huber and Roth, 1999). Reimann (1997) admitted that many factors have yet to be explored, and most studies have placed the emphasis on the 'cognitive side' without consideration for affective dimensions. Reimann's insights have been widely replicated by researchers within the field of information behaviour (*see* Chapter 4). The flexibility of Cognitive Load Theory allows a wide range of causal factors to be incorporated within a model, which can be used to provide an improved understanding of behaviour that integrates both cognitive and affective dimensions. Doyle (1963) argued that the increasing awareness of a human's incapacity to state their true need will make the evaluation of IR systems problematic. Cognitive Load Theory is a method of exploring a human's incapacity, and can attribute incapacities to causal factors.

Cognitive Load Theory cannot be used to predict an individual's behaviour. Measuring the effect an information object has on an individual's knowledge state is impossible. Cognitive Load Theory can instead be used to provide a better understanding of behavioural processes by determining the extent that causal factors affect cognitive load. An assessment of cognitive load imposed on an individual may help explain why that individual has adopted a particular behavioural pattern at a specific point in time. An individual's affective emotions can provide an insight into the development of a knowledge state over time. Alongside affective emotions, evidence of a change in an individual's knowledge state can be provided when an individual shares their interpretation of a problem state with others. Interpretation is a creative process. Such creative processes are well beyond the scope of this thesis. However, Chapter 9 proposes a framework that enables the analysis of creativity, based on the Geneplore Model (Finke, Smith, and Ward, 1996) within the domain of using relevance feedback passages to construct answers to simulated retrieval tasks.

6.5 OUTSTANDING ISSUES

- Individuals purposefully construct meaning by selectively attending to that which connects with what they already know. The ability to assimilate information is burdened by cognitive load. Both causal (intrinsic) and assessment (extraneous) factors affect cognitive load. Causal factors are the characteristics of an individual (e.g., cognitive abilities, and domain knowledge), while assessment factors are characteristics of the mental effort needed and the mental load imposed. Mental load is the portion of cognitive load that is imposed exclusively by the task and environmental demands. All these aspects must be incorporated when proposing a Cognitive Load Theory for relevance feedback (*see* Chapter 7).

- The proposed theory uses Cognitive Load Theory as a means of examining behaviour, rather than being concerned with suggesting techniques for minimising extraneous cognitive load. Cognitive Load Theory, as currently formulated, does not address the processing of highly complex information adequately, especially in language-based discursive areas such as IR. Exposure to threatening, contradictory, or complex information is likely to increase levels of anxiety. A domain expert is likely to use a schema-driven approach to information seeking. Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves without being driven by a need to reduce affective anxieties.

- Many causal factors that affect learning from instructional material also affect information behaviour. The flexibility of Cognitive Load Theory can be used to provide an improved understanding of behaviour that integrates both cognitive and affective dimensions. Although Cognitive Load Theory cannot be used to measure relevance, it could enable a greater understanding of how relevancy judgements are influenced by quantifiable *intrinsic* and *extraneous* factors.

7.0 INTRODUCTION

Cognitive load is not solely responsible for all the problems associated with the elicitation feedback. However, it can be speculated that a heavy cognitive load imposed on an individual during information seeking causes difficulties when providing relevance feedback judgements. Experimentation reported in Chapters 9-11 examine the extent these difficulties affect an individual's ability to provide feedback and how they affect the suitability/accuracy of a judgement. Experimentation requires a mechanism that enables the degree to which a feedback judgement has been inhibited by cognitive load to be evaluated. This chapter details a novel theoretical postulate that suggests how cognitive load affects information seeking. **It is postulated that when an information need progress to a point where an individual is satisfied that it has been sufficiently addressed, it becomes easier to evaluate the usefulness associated with information objects that were used to address the information need.** As discussed in Chapter 6, successful information processing requires the development of schemata (i.e., learning needs have to be met). Individuals use *heuristics* to acquire schemata (Richardson, 1986).

Simon (1969) introduced the concept of heuristics for information processing. Heuristics (or rules of thumb) direct the strategies and tactics that a user applies during information processing. Heuristics are generated to enable an individual to address a problem in a way that is compatible with their learning abilities. An individual's learning ability is limited by factors such as, short term memory, intelligence, domain knowledge, uncertainty, etc. Simon suggested that the development of new heuristics, or when a mismatch in perceptions and actual experience causes existing heuristics to be abandoned, affects cognitive load. This finding has been widely replicated (e.g., Johnson-Laird, 1983). However, nothing can be assumed about the effect cognitive load has on the modification of an individual's heuristics, other than a recognition that a modification has occurred. Cognitive load is a concept that can help to explain behaviour. It cannot be used to predict new behavioural manifestations.

The following assumptions about a Cognitive Load Theory, which attempts to explain information behaviour need to be incorporated within a theoretical model before being experimentally verified:

- Intrinsic cognitive load can be minimised by information interactions that reduce uncertainty and allow a knowledge state to be represented by enhanced schemata.
- Extraneous cognitive load can be minimised if information objects can be processed in a way that is compatible with an individual's schemata.

7.1 POSSIBLE EFFECTS OF COGNITIVE LOAD ON THE EMERGENCE OF MEANING

Figure 7.1 presents a novel theoretical model outlining how the development of a Cognitive Load Theory can help to explain information behaviour. This model shows how cognitive load can restrict an individual's ability to satisfy an information need. If cognitive load is within the bounds of mental resources, then an individual is able to process an information object without hindrance. If the information object is pertinent, then it may help the development of an individual's knowledge state and may trigger an *emergence of meaning*. If the total burden of cognitive load is too high, then an information object will not be processed in a way that enables significant knowledge state development (schemata enhancement). A failure to process an information object may cause an individual's search heuristics (problem solving strategy) to be modified. This modification may also trigger an *emergence of meaning*.

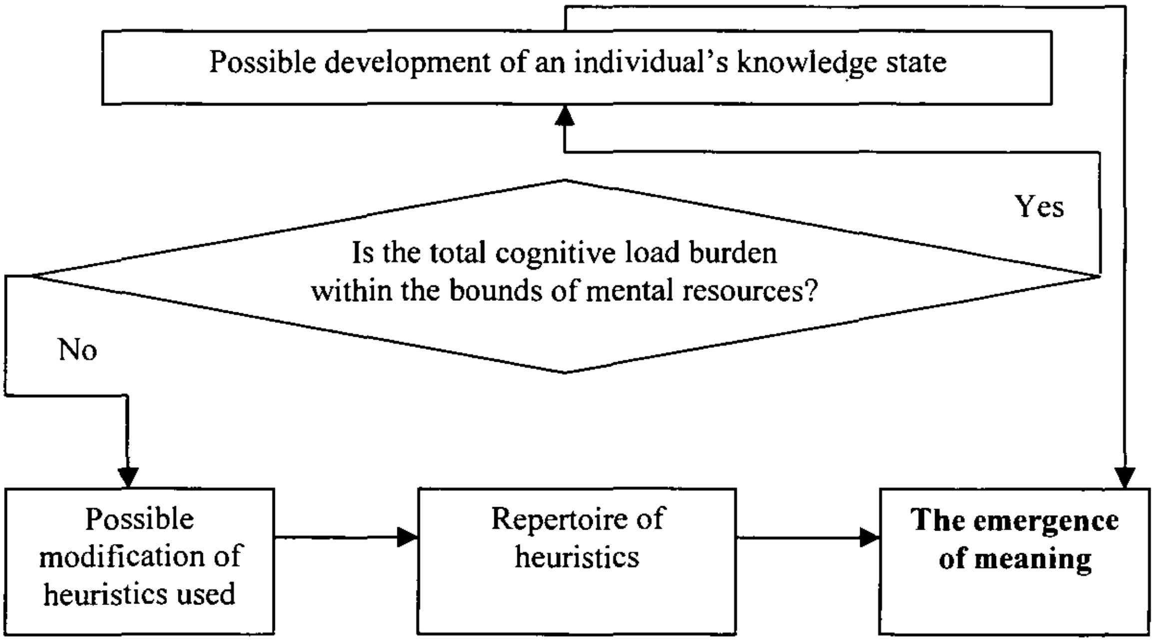


FIGURE 7.1: POSSIBLE EFFECTS OF COGNITIVE LOAD ON INFORMATION BEHAVIOUR

Brookes (1977) believed that knowledge is a structure of concepts. Knowledge structures can be transfigured by information (which is also structured). The perception of information is controlled by existing knowledge structures. Information has a largely unpredictable effect on these structures because meaning is dynamic and situational. An approach is needed that enables an understanding of how knowledge structures are transfigured and how meaning arises as a result of this transfiguration. Meaning can be better understood if it is considered to be an emergent property of a complex system of knowledge structures. Complex systems are self-organising, creative, and adaptable (Holland, 1998; Kauffman, 1995). The term 'emergence', was first coined by the English philosopher Lewes (1875). Emergence refers to the arising of novel and coherent structures, patterns, and properties during the process of self-organization in complex systems (Goldstein 1999).

Figure 7.1 shows that an emergence of meaning is central to an individual's information behaviour. If cognitive load allows the development of an individual's knowledge state, it is also likely to allow the emergence of meaning. This is because an individual's knowledge state may be transfigured in a way that provides a new insight into their information need, rather than just an accumulation of knowledge, i.e., transforming knowledge into meaning. Figure 7.1 also shows that although cognitive load may be a barrier to developing an individual's knowledge state, an emergence of meaning is still possible through modification of heuristics. If an individual applies different heuristics (problem solving strategies) to their information need, existing knowledge structures may be transfigured and new meanings may emerge. Different 'meanings' may be triggered depending on the cognitive load imposed on the individual.

Lefebvre and Letiche (1999) suggested emergent phenomena should be conceptualised as occurring on the macro level, in contrast to the micro-level components and processes out of which they arise. Applying the concept of an emerging worldview to information seeking demonstrates why modelling behaviour cannot be achieved from a purely cognitive perspective. Assigning meaning is an individualistic process that can only be observed by examining behaviour that occurs at the macro level.

De May (1977) believed that information processing is mediated by a system of concepts which, for the information processing device, are a part of its world model. Ingwersen (1992) argued that the transformation of a world model is only possible if elements of potentially relevant information are recognised.

If cognitive load is beyond mental resources, then elements of potentially relevant information will not be recognised. Transformation of a knowledge state does not necessarily produce a simple accumulation, but can be seen as a reconfiguration. Over time, the same human recipient may interpret an identical piece of information differently. Interpreting an identical piece of information differently is also possible when different levels of cognitive load are imposed. These characteristics of human information processing are also recognised characteristics of complex systems. Lefebvre and Letiche (1999) suggested that complexity is constructed from perceptual elements, each of which is experientially possible, but all of which are not simultaneously perceived.

Dent (1999) emphasised that if a single problem can be solved completely independently of everything else in the system and its environment, problem solving is an ideal strategy. However, when inter-dependencies are present, problem solving becomes less effective. Ingwersen (1992) believed that reasons for a desire for information might not be confined to problem-solving issues. Cultural goals, emotional goals, and interests must also be involved. He suggested that it becomes profoundly uncertain which elements of potentially relevant information may be of importance.

Goldstein (1999) outlined some properties of emergence that are common to both complex physical and complex artificial systems:

- **Radical novelty** - Emergents have features that cannot be observed in micro-level components.

- **Coherence or correlation** - Emergents appear as integrated wholes that tend to maintain some sense of identity over time.

- **Global or macro level loci** - Since coherence represents a correlation that spans separate components, the locus of emergent phenomena occurs at a global or macro level, in contrast to the micro-level locus of their components.
- **Dynamical**- Emergent phenomena are not pre-given wholes but arise as a complex system evolves over time.
- **Ostensive**- emergents are recognized by showing themselves, i.e., they are ostensively recognized. Because of the nature of complex systems, each ostensive showing of emergent phenomena will be different to some degree from previous ones.

How do properties of emergence manifest themselves within knowledge structures used during information processing? Fauconnier (1997) claims that language is a superficial manifestation of hidden, highly abstract, cognitive constructions. Mental spaces are the domains that discourse builds up to provide a cognitive substrate for reasoning and interfacing with the world. A sentence that appears at some stage of discourse construction will contain various grammatical devices such as descriptions regarding what new spaces are being set up (space builders), clues to which space is currently in focus (tenses and moods), descriptions that identify existing elements, and descriptions that introduce new elements into spaces. An extensive analysis of discourse has resulted in an improved understanding of how meaning emerges from information. Fauconnier described conceptual networks (knowledge structures) as being intricately structured by analogical and metaphorical mappings. These mappings play a key role in the construction of meaning and its evolution:

- **Analogy and Schema Induction**- Recognition of generic level features that can be related to existing knowledge structures.
- **Categorisation and New Conceptual Structure**- Classification of a novel category that will fit generic conceptual specifications.
- **Naming and Projected Structure**- Unique attributes are identified and a novel conceptual schema is developed.
- **Blending and Conceptual Integration**- Blending and integration of conceptual schemes resulting in emergent properties.

- **Motivated Polysemy-** Instead of blending, schemata may become increasingly distinguished but without loosing analogical links.

- **Divergence and Extinction-** Instead of blending, conceptual links may disappear. Cognitive importance is not reduced by the fact we cease to be consciously aware.

Where does the process of emergence manifest itself during information seeking and what effect does cognitive load have? Ingwersen (1992) outlined five cognitive structures that are interweaved (layered) within the knowledge state of an individual.

The cognitive model layer encompasses all other layers. The 'work space' consists of interests, conceptual knowledge, tasks, and information seeking behaviour. This 'work space' comprises of an individual's cognitive abilities. Deficiencies in this area could cause a high cognitive load. The actual state of knowledge layer represents consciously available knowledge at search time (working memory). The problem space layer consists of the recognised problem situation. The problem space layer could also cause a high cognitive load to be imposed if the problem situation is complex. The state of uncertainty layer represents the desire for information.

Table 7.1 outlines a novel framework that attempts to show how meaning emerges from information. The purpose of this model is to demonstrate that the emergence of meaning is a phenomenon that cannot be modelled easily. Furthermore, this model implies that explaining information behaviour by attempting to isolate the cause of changes in meaning assigned to information is a futile pursuit.

ASK *Belkin (1977)* **Processing Stages** *De Mey (1977)*	**Analogical and Metaphorical Mappings** *Fauconnier (1997)*	**Discourse Construction** *Fauconnier (1997)*	**Cognitive Structures** *Ingwersen (1992)*	**Emergent Properties** *Goldstein (1999)*
ASK Recognition Stage (1)	N/A	N/A	State of Uncertainty	***Emergents arise as a complex system evolves*** ***Each ostensive showing will be different from previous ones***
Monadic Stage (2)	Analogy and Schema Induction	N/A	Problem Space	
Structural Stage (3)	Categorisation and New Conceptual Structure	Space Builders	Actual State of Knowledge	***Emergents have global or macro level loci they, cannot be predicted during these stages***
Contextual Stage (4)	Naming and Projected Structure	Descriptions that Identify Existing Elements, **AND** Tenses and Moods (Focus)	Work Space	
Cognitive or Epistemic Stage (5)	**EITHER** Blending and Conceptual Integration, **OR** Motivated Polysemy, **OR** Divergence and Extinction	Descriptions which Introduce New Elements	Cognitive Model	***Emergents can exhibit radical novelty*** ***Emergents appear as integrated wholes that maintain identity***

TABLE 7.1: EMERGENCE OF MEANING

7.1.1 STAGE ONE

The first stage of the framework is termed the ASK stage. According to the cognitive viewpoint, an ASK can be considered as being functionally equivalent to a state of uncertainty because they both represent a cognitive breakdown situation (Ingwersen, 1992). A cognitive breakdown situation is recognition of a lack of knowledge and is caused by a series of complex interactions. Therefore, an ASK can be conceptualised as a novel emergent property of interacting knowledge states.

When examining information interactions, we are concerned primarily with: how an information need is developed; how it is represented; how both are affected by the technical and conceptual tools available; what problem-solving strategies are adopted; and what reasons there are for adopting these strategies. If a *sufficient* number of variables, constraints, and interactions are identified, it is theoretically possible to mathematically model any phenomenon. Unfortunately, when examining cognitive processes, it is impossible to discover when a sufficient number of properties have been found. This is because humans do not perform exhaustive searches of a problem space, even if the problem is simple enough to make it feasible (Richardson, 1986; Johnson-Laird, 1983). Humans use heuristics instead. An ASK is a product of evolving/interacting knowledge states. An ASK is an ostensive occurrence that cannot be predicted. Understanding human cognition, when the act of problem solving seems to escape logical analysis, can result in a pervasive tendency to overestimate the value of human judgement (Bechtel and Richardson, 1993). A human judgement is inhibited by cognitive load. Therefore, the use of self-report during information seeking is not an entirely reliable way of gaining a full understanding of information behaviour. Requiring an individual to self-report decisions made during information seeking is fortunately not the only way of gaining an insight into information behaviour. Behavioural patterns and easily quantifiable variables (such as uncertainty) can also be analysed (*see* Chapter 9).

7.1.2 STAGES TWO AND THREE

Examining the monadic and structural stages of information processing has been achieved within the best-match (physical) paradigm. It has been discovered that an information need cannot be represented using logical constructs. Within the monadic and structural stages, artificial static constructs are the only way of representing an information need. Relevance feedback in IR currently operates within the monadic and structural stages of information processing. Without consideration of the cognitive load concept, relevance feedback cannot operate at the contextual and epistemic levels of information processing.

Belkin, Oddy, and Brookes (Republished, 1997) claimed that relevance feedback seems more responsive to the information need issue, in that it utilises user relevance judgements, and therefore is not solely dependent on the characteristics of the document collection. However, relevance feedback remains within the best-match (physical) paradigm, since it assumes that the eventual query formulation is equivalent to the ideal document. Fauconnier's models (1997) show that during the monadic and structural stages, nothing more than matching and categorisation is performed. Ingwersen's model shows that nothing more than static tasks and static knowledge structures are operated upon. These environmental conditions are ideal for comparative relevance feedback testing using traditional objective evaluation measures.

Much relevance feedback research has used artificial pre-judged relevance judgements, with a lack of knowledge on how human relevance judgements occur and develop in an interactive IR environment. The evaluation of relevance feedback mechanisms has only occurred at monadic and structural levels. Outside the laboratory environment, user-based relevance feedback techniques have been of limited success. The majority of Web search engines and online database providers have abandoned attempts at incorporating a user relevance feedback system. An evaluation procedure is needed at the epistemic level in order to understand why relevance feedback is not effective outside the laboratory environment (*see* Section 7.2).

7.1.3 STAGES FOUR AND FIVE

The user-centred and cognitive approaches to IR attempt to operate within the contextual and epistemic stages of information processing. Fauconnier's models show that these stages are where meaning is derived and knowledge structures are modified. Ingwersen's model (1992) shows that changes in behavioural and cognitive process need to be modelled if interaction is to be understood. Table 7.1 (on page 105) reveals that because emergents have global or macro level loci, they cannot be observed during monadic, structural, or contextual stages. They can only arise at the ASK recognition stage or at the epistemic stage. Meaning seems to exhibit the same properties of an emerging phenomenon. Without considering the extent to which potentially useful information is recognised or re-configured by a user, relevance feedback cannot convey meaning. Figure 7.3 (on page 110) shows that the process of relevance feedback is influenced by interactions that occur at all levels of information processing. When evaluating the effect cognitive load has on relevance feedback, a technique is needed that does not isolate information processing stages.

Philosophers have argued that the separation of meaning from observation is not feasible. Separating information processing stages is not feasible because it prevents the emergence of meaning from information. It can also be argued that the emergence of meaning always requires novel conceptual insights. Buchdahl (1969) noted that the philosopher Kant believed that we should not think of ideas as aiming at a mere increase in the probability of the resultant explanation; ideas could never begin if the constructive approach were not postulated from the start, governing the form of the idea itself. Kant's proposals, originally made in 1781, are almost identical to Kuhn's (1996, p. 97) "paradigms provide all phenomena except anomalies with a theory-determined place in an individual's field of vision", and Feyerabend's (1993, p. 61) "the content of a concept is determined also by the way in which it is related to perception". Feyerabend (1993) suggested that counterinduction is an essential part of discovery. Ideological ingredients of our knowledge / observations are discovered with the help of theories which are refuted by them[24].

[24] This should not be confused with Popper's (1959) proposal of the *importance of falsification*. The imperfections of existing data-theory fit can define new meanings. Falsification does not always occur because of the emergence of an anomaly.

7.2 EVALUATING RELEVANCE FEEDBACK

As previously discussed in Chapter 6, cognitive load has both intrinsic and extraneous components. One of the aims of experimentation reported in Chapters 9, 10, and 11 is to identify which information seeking variables can be used to predict the affect of intrinsic and extraneous cognitive load. However, before this can be achieved, the effect that intrinsic and extraneous cognitive load has on relevance feedback needs to be postulated.

Intrinsic components comprise of an individual's cognitive capabilities associated with addressing their information need, e.g., domain knowledge, uncertainty, and expectations are easily recognisable intrinsic components. If an individual has well developed knowledge schemata, more intrinsic elements can integrated within an individual's working memory at search time. For example, good knowledge schemata will encompass more domain knowledge, have a lucid idea of what information is missing/required, and can predict the type/nature of information required (expectations).

Extraneous components comprise of the cognitive difficulties associated with processing potentially pertinent information objects. For example, cognitive style, learning and comprehension abilities, and intelligence, are easily recognisable extraneous components. These types of individual attributes are fixed. Depending on the nature of information that needs to be processed, the extraneous cognitive load imposed is representative of an individual's ability to process such information.

Figure 7.2 outlines a novel postulate, discussed further in Chapter 12, that shows that intrinsic and extraneous cognitive load has the potential to affects an individual's heuristics (search strategies) in different ways. It is hypothesised that the intrinsic cognitive load imposed on an individual limits the type of heuristics deemed suitable by that individual for addressing their information need. Depending on an individual's capabilities, the most suitable repertoire of heuristics will be selected by them. It is hypothesised that the extraneous cognitive load imposed affects the way in which heuristics are modified. Modification of heuristics is essential to enable information objects to be processed in a way that is compatible with an individual's existing knowledge schemata.

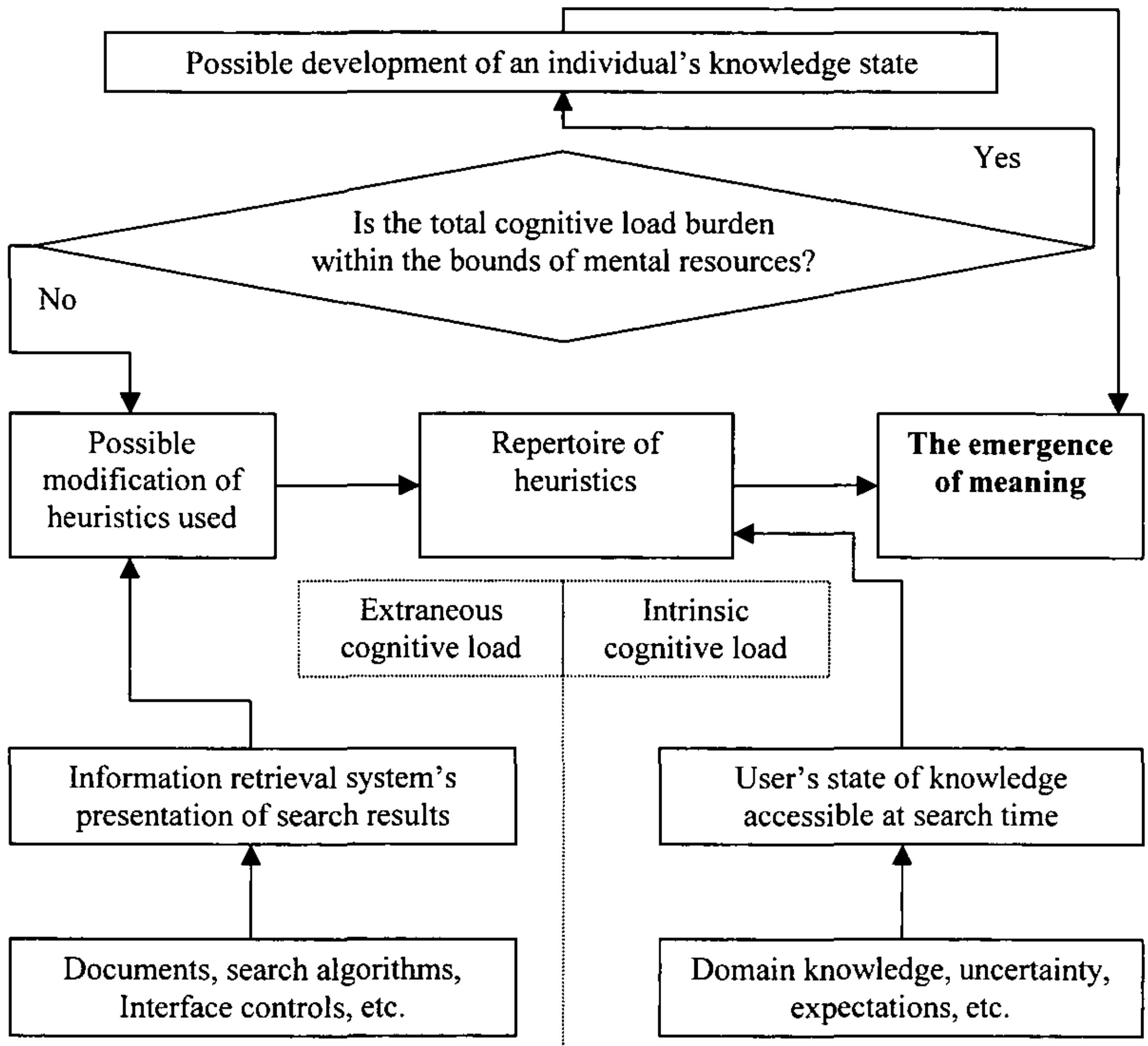


FIGURE 7.2: POSSIBLE EFFECTS OF INTRINSIC AND EXTRANEOUS COGNITIVE LOAD ON INFORMATION BEHAVIOUR

This thesis will not attempt to develop a Cognitive Load Theory that attempts to explain all aspects of information behaviour. Developing such a theory cannot be achieved without a better understand of the individual processes involved during information seeking. This thesis is concerned with the issue of relevance feedback. Figure 7.3 (on page 112) shows that the process of relevance feedback is influenced by interactions that occur at all levels of information processing. When evaluating the effect cognitive load has on relevance feedback, a technique is needed that does not isolate information processing stages.

Within this thesis, cognitive load for relevance feedback is defined as the intrinsic and extraneous difficulty associated with providing useful / accurate feedback judgements during a search task. This thesis hypothesises that more useful / accurate feedback judgements will be made if both an individual's intrinsic cognitive load and extraneous cognitive load is low. Figure 7.3 details a novel theoretical postulate that proposes how cognitive load affects relevance feedback.

As previously discussed, intrinsic load is not affected by the information objects retrieved by an IR system. Figure 7.3 shows that intrinsic load is imposed at the epistemic (or cognitive) stage of information processing. Intrinsic factors can be conceptualised as components of an individual's 'anomalous state of knowledge' (Belkin, 1977), or 'state of uncertainty' (Ingwersen, 1992). Essentially, intrinsic factors are components of an individual's knowledge state, which is comprised of schemata (Cooper, 1990), which are accessed using heuristics (Simon, 1969). The state of an individual's knowledge governs the amount of cognitive load imposed on the individual.

Sophisticated knowledge states can be held more easily in working memory. Less sophisticated knowledge states cannot be held entirely within working memory so a higher cognitive load is imposed. These principles are identical to the ones proposed by all researchers who apply Cognitive Load Theory to their various research fields (*see* Section 6.2). No new assumptions about an individual's cognitive architecture have been made.

As previously discussed, extraneous load is affected by the information objects retrieved by an IR system. Figure 7.3 shows that extraneous load is imposed at the contextual (or situational) stage of information processing. Extraneous factors can be conceptualised as components of an individuals 'work space' (Ingwersen, 1992) or 'naming and projected structure stage' (Fauconnier, 1997). Extraneous load is governed by the suitability of an individual's heuristics at processing information objects. More suitable heuristics enable information to be processed in a way which is more compatible with existing knowledge schemata.

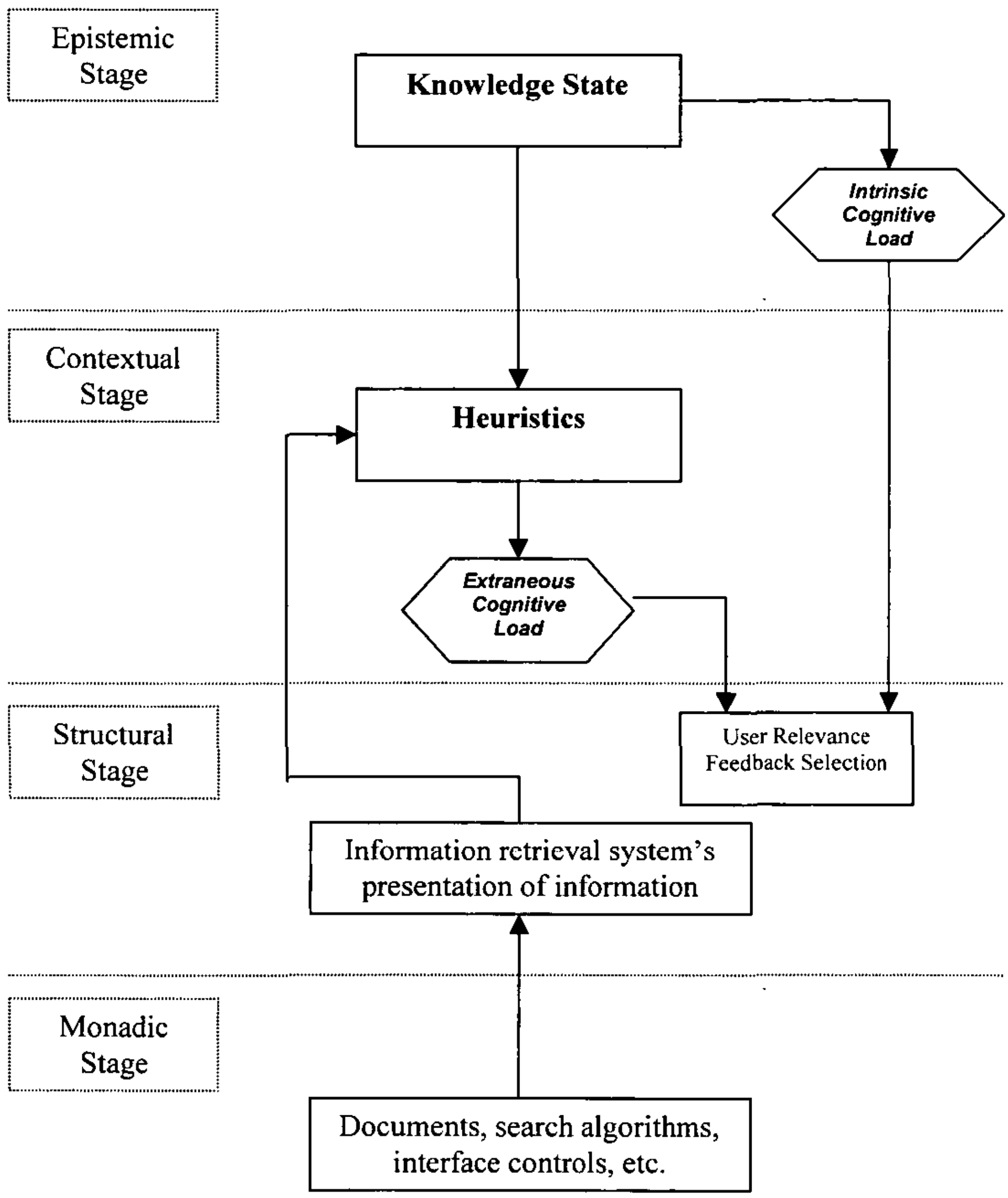


FIGURE 7.3: EVALUATING RELEVANCE FEEDBACK

Figure 7.3 shows that although relevance feedback judgements are provided at the structural stage (by identifying relevant semantic elements of documents), the judgement process is affected by both intrinsic and extraneous cognitive load.

When evaluating relevance feedback systems using human participants, a wide range of situational and behavioural variables can be examined. Analysing the interaction between variables after a search task is completed is not psychologically realistic. The behavioural context in which relevancy judgements are generated cannot be revealed without an understanding of the cognitive load imposed. Understanding cognitive load may help to explain why individuals adopt a specific behavioural style. Cognitive load allows the psychological context of behaviour to be examined. It does not allow an insight into the process of how an individual selects, applies, and modifies heuristics. Understanding information behaviour at the cognitive (epistemic) level may not be possible. The only reliable way of examining information behaviour is to find explanations for observable behavioural patterns.

7.3 THEORETICAL POSTULATE: CAN MEASURES OF COMPARISON BE USED TO ASSESS COGNITIVE LOAD?

A technique is needed that allows a number of behavioural variables to be considered for inclusion within mathematical models of intrinsic and extraneous cognitive load. Regression analysis allows specific predictions to be made from an independent variable (IV) about the dependent variable (DV). Multiple regression analysis allows more than one independent variable to be used to predict the dependent variable and so improve the accuracy of the prediction. The aim is to generate a mathematical model that enables the accurate prediction of the dependent variable. Obviously, before it is possible to identify which behavioural variables (DVs) contribute to the different intrinsic and extraneous cognitive load accounts (IVs), a way of quantifying cognitive load is required.

This thesis hypothesises that cognitive load is a barrier to providing useful relevancy judgements. A mechanism is required that enables the degree to which a judgement has been inhibited by cognitive load to be evaluated. The mechanism postulated below is inherent within the dynamic nature of providing relevancy judgements. A new measurement tool need not be invented.

A relevancy judgement is a dynamic phenomenon. Judgements are provided at a specific point in time and will fluctuate as an information need develops. The cognitive load imposed on an individual is also dynamic. Cognitive load fluctuates throughout all learning (or problem solving) processes. It is postulated that when a search has progressed to a point where an individual is satisfied that their information need has been sufficiently addressed, it becomes easier to evaluate the usefulness associated with information objects that were used to address the information need. If an individual is asked to re-evaluate the usefulness of an information object used, a comparison can be drawn between previous relevancy assessments made for relevance feedback purposes, and the re-evaluated assessment. The difference between these judgements is likely to reveal a discrepancy. This discrepancy, or degree of error in judgement accuracy, can be used as a measure of comparison (*mc*). This measure of comparison can be used to estimate the degree of cognitive load imposed on an individual when the initial relevance feedback judgement was submitted during the search process for a specific information object (*j*) (*see* Equation 7.1).

$$mc_j = \frac{cl}{bcl} = \frac{|ij_j - pj_j|}{\left(\frac{pj_j}{tfe}\right)}$$

Where

cl	= Cognitive Load
bcl	= Baseline Cognitive Load
ij_j	= Initial Judgement assigned to information object j
pj_j	= Post-task Judgement assigned to information object j
tfe	= Task Fulfilment Evaluation

EQUATION 7.1

A baseline cognitive load measure (*bcl*) would enable the accuracy of a post-task judgement to be assessed. A task fulfilment evaluation enables the extent to which an individual is satisfied that their information need has been addressed to be evaluated. This has the effect of normalising the accuracy of a post-task judgement (pj_j) by establishing a baseline measure (*bcl*). For example, if an individual is not satisfied that they have completed a task satisfactorily, the post-task judgement (pj_j) is not considered to be more accurate that the initial judgement (ij_j).

Equation 7.1 is a theoretical postulate of how it **may** be possible to measure the cognitive load imposed on an individual during the process of relevance feedback. Experimentation reported in Chapter 8 explores if the above postulate can be successfully implemented. It also scrutinizes the assumption that post-task judgements are more accurate than an initial judgement made for relevance feedback purposes.

The next step is to suggest how the development intrinsic and extraneous measures of comparison are possible. The development of intrinsic and extraneous measures of comparison enables prospective contributors to a model of cognitive load to be evaluated. When identifying intrinsic contributors, the DV is the intrinsic measure of comparison (*imc*). The extraneous measure of comparison (*emc*) is used as the DV when identifying extraneous contributors. This next step cannot be taken before it is established how cognitive load can be assessed during information seeking (*see* Chapter 8).

7.4 OUTSTANDING ISSUES

- Human judgement is inhibited by cognitive load. Therefore, the use of self-report during information seeking is not a reliable way of gaining a full understanding of information behaviour. Heuristics are generated to enable an individual to address a problem in a way that is compatible with their learning abilities. An individual's learning ability is limited by, short term memory, intelligence, domain knowledge, uncertainty, etc. If an information object is pertinent, then it may help the development of an individual's knowledge state and may trigger an *emergence of meaning*. A failure to process an information object may cause an individuals search heuristics (problem solving strategy) to be modified. This modification may also trigger an *emergence of meaning.*

- Attempting to explain information behaviour by attempting to isolate the cause of changes in meaning assigned to information, is a futile pursuit. Without consideration of the cognitive load concept, relevance feedback cannot operate at the contextual and epistemic levels of information processing. Understanding cognitive load may help to explain why individuals adopt a specific behavioural style. Cognitive load does not allow an insight into the process of how an individual selects, applies, and modifies heuristics.

- Intrinsic factors are components of an individual's knowledge state, which is comprised of schemata, which are accessed using heuristics. The state of an individual's knowledge governs the amount of cognitive load imposed on the individual. Sophisticated knowledge states can be held more easily in working memory. Less sophisticated knowledge states cannot be held entirely within working memory, so a higher cognitive load is imposed. Extraneous load is governed by the suitability of an individual's heuristics at processing information objects. More suitable heuristics enable information to be processed in a way which is more compatible with existing knowledge schemata.

- A mechanism is required that enables the degree to which a judgement has been inhibited by cognitive load to be evaluated. A relevancy judgement is a dynamic phenomenon. Judgements are provided at a specific point in time and will fluctuate as an information need develops. The cognitive load imposed on an individual is also dynamic. Cognitive load fluctuates throughout all learning (or problem solving) processes. It is postulated that when an information need progress to a point where an individual is satisfied that it has been sufficiently addressed, it becomes easier to evaluate the relevancy associated with information objects that were used to address the information need. Experimentation is needed to scrutinize the assumption that post-task judgements are more accurate, due to lower levels of cognitive load, than an initial judgement made for relevance feedback purposes (*see* Chapter 8).

8.0 INTRODUCTION

This chapter reports on two small-scale exploratory studies. The results of these studies were used to develop the main experimental procedure detailed within Chapters 9-12 of this thesis. Exploratory Study 2 explores the fundamental premise that cognitive load is a measurable phenomenon. Exploratory Study 3 provides an insight into the behavioural patterns associated with providing passage relevance feedback. Additionally, Exploratory Study 3 attempts to identify behavioural variables that could *potentially* be incorporated within a cognitive load model of feedback. The aim of these experiments is to find evidence to enable the development of intrinsic (*imc*) and extraneous (*emc*) measures of comparison, which can be used to enable contributors to a model of cognitive load to be identified.

8.1 EXPLORATORY STUDY 2: EVIDENCE FOR THE DEVELOPMENT OF COGNITIVE LOAD THEORY

The extent that cognitive load might have compromised the usefulness / usability of relevance feedback systems needs to be measurable. To enable measurement, comparative measures need to be developed. The postulate that post-task judgements are more accurate, due to lower levels of cognitive load than an initial judgement made for relevance feedback purposes, needs to be tested (*see* Chapter 7). The aims and objectives of Exploratory Study 2 were as follows:

Aims:

1. To determine whether variations in initial and post task usefulness judgements are indicative of the cognitive load imposed on an individual.
2. To determine whether intrinsic and extraneous cognitive load accounts can be differentiated.

Objectives needed to fulfil Aim 1:

A. Implement a procedure that allows both initial and post task usefulness judgements to be assigned within the context of satisfying a simulated information need.

B. Evaluate the extent to which usefulness measures are suitable for the measurement of cognitive load.

Objective needed to fulfil Aim 2:

C. Evaluate whether the process of 'making notes' during information seeking can reveal how extraneous cognitive load affects usefulness judgements.

8.1.1 METHOD

Microsoft Visual Studio 6.0 was used to write the software required for this experimental procedure (*see* Appendix B for screenshots). A 'Web browser' style interface was developed. This type of interface was chosen on the basis that all participants selected for this experiment were familiar with navigating Web pages. Information presented to experiment participants was cached from the Web version of the Encyclopaedia Britannica (permission was obtained). Standardised instructions were presented to participants alongside a questionnaire that gathered background data (*see* Appendix B).

Participants were required to perform two simulated tasks from a pool of ten. Participants were able to select which tasks they wanted to perform. This procedure attempted to minimise potential problems associated with participants undertaking tasks that they had no interest in. Encyclopaedia Britannica had provided a list of fifty of the most popular topic areas searched for in August 2000. This list was used to help generate the pool of ten tasks. Participants volunteered from a pool of final year undergraduate students in the Department of Information Science at Loughborough University. Experimentation took place during November 2000. An email was sent out inviting them to volunteer for this study. The email included the list of the fifty topic areas provided by the Encyclopaedia Britannica, although students were not informed of the source. Volunteers voted for the ten topic areas they were most interested in. These votes were collated and the top ten topic areas were used to generate the simulated tasks.

Simulated tasks were developed so that they were intricate / vague in nature and often required a multipredicate hypothesis to be tested. The intricate / vague nature of the tasks encouraged a participant to continually re-assess the requirements of the task. This more accurately models the true nature of an information need. Requiring a multipredicate hypothesis to be tested meant that multiple information sources had to be used to enable the task to be addressed successfully. This is likely to induce more cognitive load. Section 9.5 further justifies the rationale behind using simulated tasks.

Six documents (HTML pages with no hyperlinks) were presented to users in a random order. Each participant was asked to evaluate the usefulness of the document. This evaluation was performed using a magnitude scale (*see* Appendix B). Usefulness judgements were based on a participant's perception of their assigned work task. It was anticipated that each participant would interpret the work task differently to simulate a 'cognitive breakdown situation'. It was important that judgements related to the participant's mental state at the time of being presented with a Web page. Participants were informed that usefulness judgements should be 'their own opinion of how important the information is'.

For each task, pages were cached from the appropriate subject section of the Encyclopaedia Britannica. Therefore, the six documents presented for each task were all topically very relevant. Each subject section of the online encyclopaedia has an introduction page. The first six linked documents on the introduction page were the ones cached. The introduction page was not cached. This forced participants to ascertain the conceptual links between the six documents for themselves (inducing cognitive load). Therefore, it was not always immediately obvious why all the documents were relevant. If a linked document had its own links, then these links were deactivated. Participants were informed that they could revisit a page if they wanted to, but had to provide a usefulness judgement first time round even if they found it hard to do so.

Participants were actively encouraged to make notes on paper throughout the experiment. Although this introduced an artificial element to the experimental procedure, it was anticipated that the notes made would reveal valuable insights into the nature of cognitive load.

The documents presented to volunteers varied in length from 500 - 2000 words. Participants were encouraged to browse documents rather than read them thoroughly. Most participants spent, on average, four minutes per document. Information was generally complex, very concise, and often intriguing. All the participants claimed not to have problems with comprehension but often found the simulated work tasks challenging.

After all documents were shown, participants were asked to orally suggest a solution to the work task. It was envisaged that if no evidence of a basic understanding was evident, a participant would not be asked to perform the second stage of the experiment. However, this situation did not occur.

The second stage of the work task involved participants re-evaluating their usefulness judgements. The six documents were presented, again at random, for re-evaluation. The scale used to re-evaluate the documents was the same one used to obtain the initial usefulness judgement. Initial usefulness judgements were related to the participant's mental state at the time of being presented with a Web page. Re-evaluated judgements were post-task judgements. Participants were informed that post-task judgements should be 'their own opinion of how important the information was'. The same criterion was given for both the initial usefulness judgement and the post-task judgement. If participants had made notes during the task they were also asked to re-evaluate those by putting either a cross, tick, or question mark alongside them.

When a document was presented back for re-evaluation, a randomly selected half of the experiment participants were not shown their initial judgement. The other half were presented with their initial judgement, which could be used as a frame of reference. Although no time limit was set, the average time to complete two tasks was 55 minutes.

8.1.2 RESULTS

Background Data

Experimentation took place during November 2000. Twenty volunteers were selected for experimentation: 13 males, and seven females. Participants were selected on a 'first come first served' basis from a pool of 32 volunteers who were final year undergraduate students in the Department of Information Science at Loughborough University. All participants regularly used IR systems (mainly Web search engines). None of the participants were familiar with providing feedback judgements. Results showed that differences in experimental performance were not related to sex, age, or degree course.

Task Success Rate

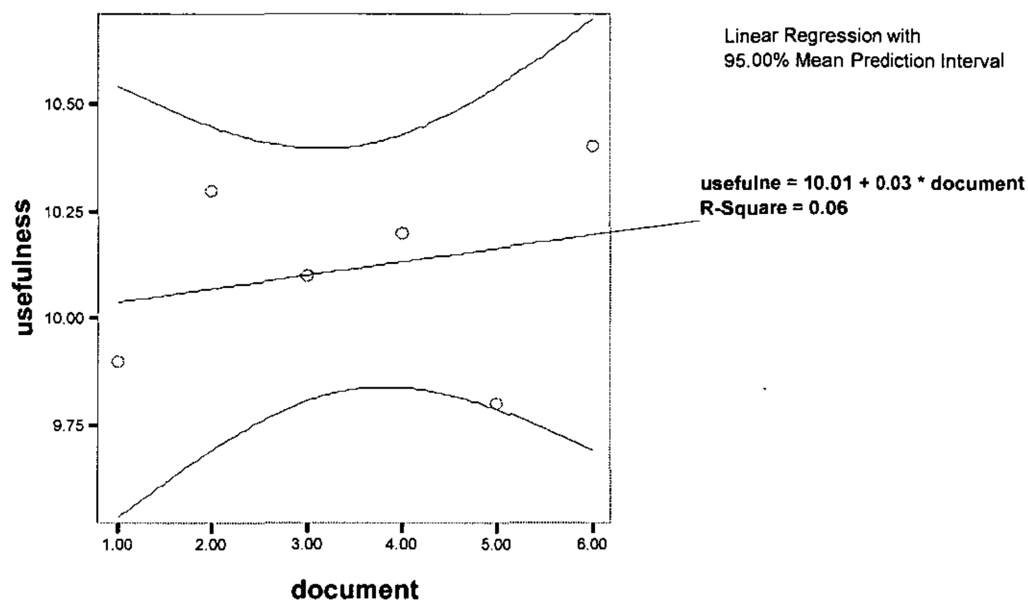
Tasks were generally well received. Many participants (70%) suggested that they were interesting. Table 8.1 shows the task success rate. The Table indicates that 90% had a good or satisfactory understanding of the task. Categorisation of a participant's task success was based on an oral answer provided after the six documents were initially evaluated and before the re-evaluation procedure. It was important to set the task difficulty at a level that would induce cognitive load. Although categorising answers in this way is by no means a precise method, these results tentatively suggest that the vast majority of participants (95%) were able to attempt the task, and that the tasks themselves were not too simplistic, as 68% did not resolve the task fully.

PROBLEM SITUATION	TASK RESOLUTION	PARTICIPANTS
Good understanding	Complete	13
Good understanding	Partial	4
Satisfactory understanding	Partial	19
Basic understanding	Partial	2
Basic understanding	No evidence	2
No understanding	No evidence	0
TOTAL		20x(2 Tasks) = 40

TABLE 8.1: TASK SUCCESS RATE

Initial usefulness judgement

All participants claimed they understood the principle of assigning usefulness judgements. Figure 8.1 shows that the usefulness of documents did not significantly improve or worsen as more documents were presented. Both a linear regression and a Pearson's correlation were used to verify this. This finding suggests that the usefulness of a document can be assessed on an individual basis, independently of other documents. Changes in domain knowledge do not affect usefulness judgements.



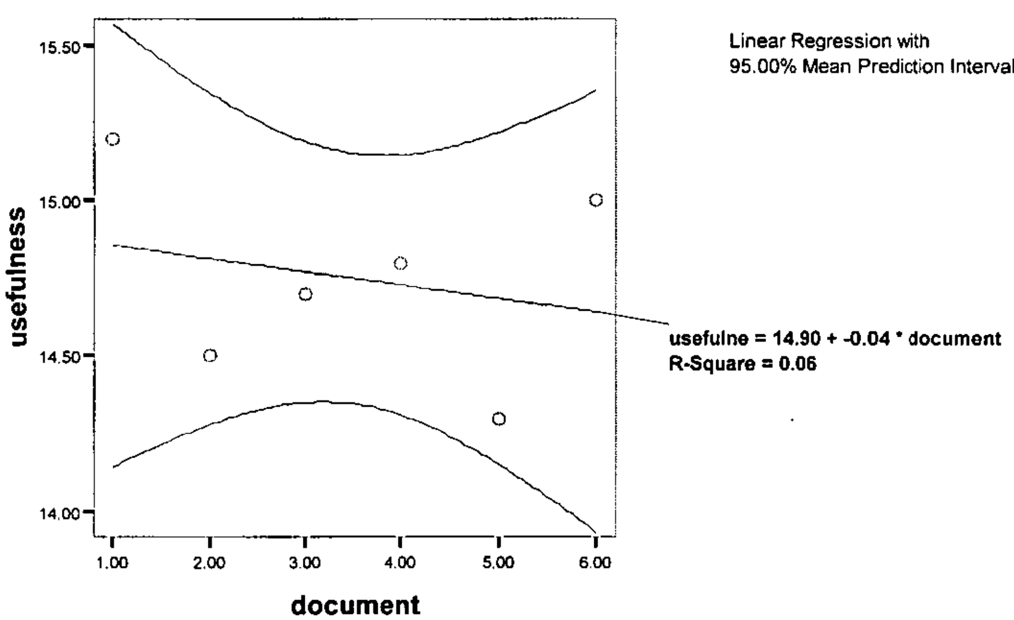
Correlations

		DOCUMENT	USEFULNE
DOCUMENT	Pearson Correlation	1.000	.254
	Sig. (2-tailed)	.	.627
	N	6	6
USEFULNE	Pearson Correlation	.254	1.000
	Sig. (2-tailed)	.627	.
	N	6	6

FIGURE 8.1: INITIAL USEFULNESS JUDGEMENT

Post task (re-evaluation) usefulness judgement

The same testing procedure was used to determine if post-task usefulness judgements were reliable. Figure 8.2 shows that no confounding variables were found that could be associated with the idea that post task judgements may improve or worsen as more documents are presented for re-evaluation. Both initial and post-task usefulness judgement are therefore a *potentially* suitable surrogate measures of cognitive load as they are not affected by the natural progression of the information seeking process (*see* Chapter 12).



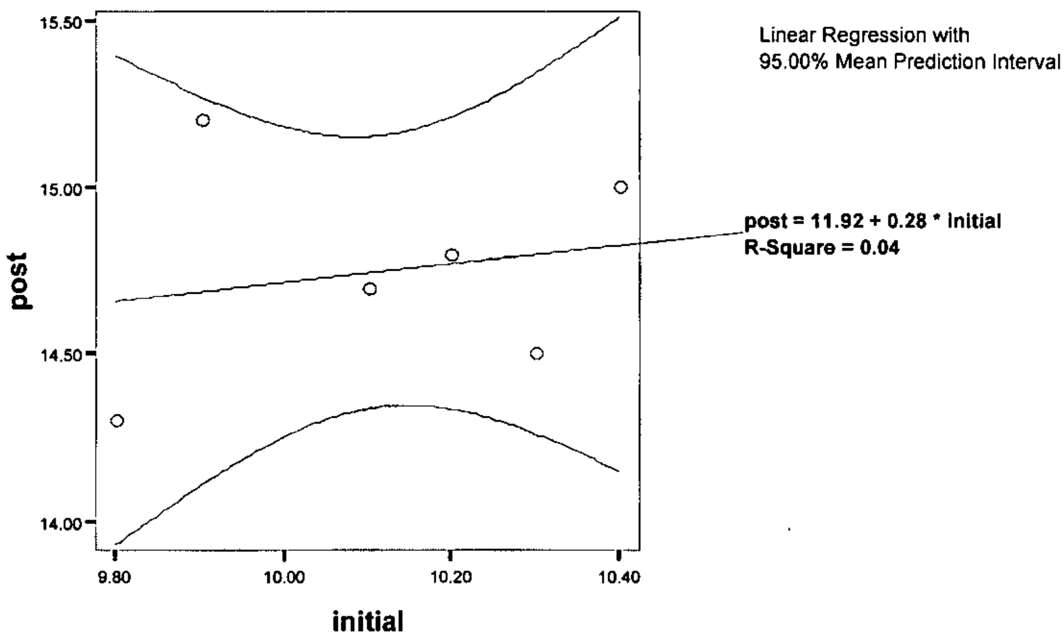
Correlations

		DOCUMENT	USEFULNE
DOCUMENT	Pearson Correlation	1.000	-.245
	Sig. (2-tailed)	.	.640
	N	6	6
USEFULNE	Pearson Correlation	-.245	1.000
	Sig. (2-tailed)	.640	.
	N	6	6

FIGURE 8.2: POST TASK USEFULNESS JUDGEMENT

Initial usefulness judgement v Post task (re-evaluation) usefulness judgement

The final reliability check required is to investigate whether an initial judgement influences a post-task judgement. Figure 8.3 indicates that although a slight positive correlation exists, it is not significant. This slight correlation may be indicative of a tendency by participants to prove that they increasingly recognised that all the documents presented were topically relevant. Some participants were keen to point this out.



Correlations

		INITIAL	POST
INITIAL	Pearson Correlation	1.000	.198
	Sig. (2-tailed)	.	.707
	N	6	6
POST	Pearson Correlation	.198	1.000
	Sig. (2-tailed)	.707	.
	N	6	6

FIGURE 8.3: INITIAL V POST TASK USEFULNESS JUDGEMENT

Re-evaluation variance with or without a frame of reference?

When a document was presented back for re-evaluation, half of the experiment participants were not shown their initial judgement. The other half were presented with their initial judgement, which could be used as a frame of reference. Table 8.2 shows that when the initial judgement was not shown, the average variance between initial and post judgements was 23%. When the initial judgement was shown, the average variance was 25%. This finding shows that no significant differences exist. The use of a frame of reference does not affect the re-evaluation of usefulness. Usefulness can therefore be used as a variable without requiring the use of comparative judgements.

PROBLEM SITUATION	PARTICIP.	VARIANCE
No frame of reference	10	23.020%
Initial judgement shown	10	24.645%
TOTAL	20	Mean = 23.833%

TABLE 8.2: USING A FRAME OF REFERENCE FOR RE-EVALUATION

What evidence is there that variations in initial and post-task usefulness judgements are indicative of the cognitive load imposed on an individual?

Table 8.3 shows that when a participant was able to demonstrate a good understanding, the average variance of a usefulness re-evaluation judgement was 30%. If a participant had a satisfactory or basic understanding, average variance of a usefulness re-evaluation judgement was 19%. These findings suggest that if a good understanding was achieved, then it is easier to identify which documents were useful when a task had been completed. It can be assumed that an individual with a good understanding has developed better knowledge schemata. The development of better schemata reduces cognitive load. Therefore, it can be tentatively suggested that participants who achieved a satisfactory or basic understanding had more cognitive load, so found it harder to identify which documents were useful when a task had been completed. This finding supports the need for a baseline cognitive load measure to be established (*see* Section 7.3). Task performance needs to be considered when developing a Cognitive Load Theory.

PROBLEM SITUATION	PARTICIP.	VARIANCE
Good understanding	17	30.144%
Satisfactory or basic understanding	23	19.168%
TOTAL	40	Mean = 23.833%

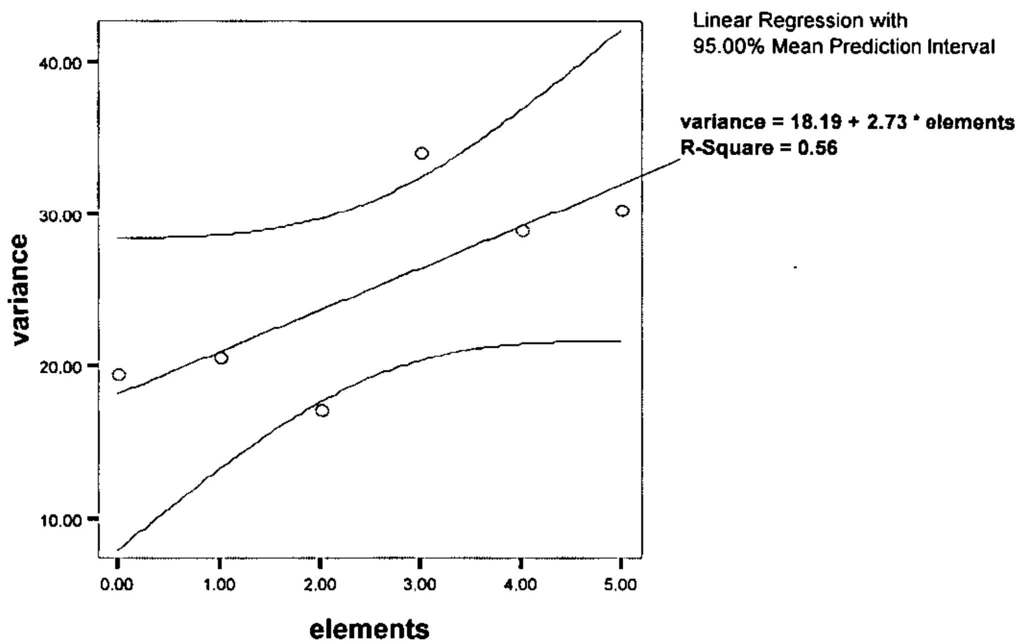
TABLE 8.3: JUDGEMENT VARIANCE AND TASK SUCCESS RATE

Is note making indicative of the cognitive load imposed on an individual?

Table 8.4 shows how the notes made by participants, for each document browsed, were categorised. Most participants (80%) reported that they found that making notes helped them to manage tasks. The number of conceptual elements (i.e,. 'things talked about') per document was recorded. If there was doubt, then a participant was asked if elements (i.e. concepts / ideas) should be regarded as being separate (or independent) from each other. Figure 8.4 shows that there is a significant positive correlation (1 tailed at 0.05 level) between the number of elements and variance. This is a crucial finding. The greater the number of elements incorporated within notes, the higher the degree of variance was found. Individuals 'talk about' elements within notes because they cannot incorporate these fully within working memory. Poorly developed knowledge schemata cannot be fully integrated within working memory (*see* Section 6.1).

ELEMENTS NOTED PER DOCUMENT	FREQUENCY	VARIANCE
5 elements +	13	30.213
4 elements	37	28.865
3 elements	45	33.945
2 elements	51	17.125
1 element	30	20.547
No notes made	64	19.404
TOTAL	240	Mean = 23.833%

TABLE 8.4: ELEMENT ANALYSIS



Correlations

		ELEMENTS	VARIANCE
ELEMENTS	Pearson Correlation	1.000	.747*
	Sig. (1-tailed)	.	.044
	N	6	6
VARIANCE	Pearson Correlation	.747*	1.000
	Sig. (1-tailed)	.044	.
	N	6	6

*. Correlation is significant at the 0.05 level (1-tailed).

FIGURE 8.4: ELEMENTS V VARIANCE

Results suggest that the accuracy associated with an initial usefulness judgement is burdened by cognitive load. Evidence has been found that shows cognitive load to be a measurable phenomenon. An increase in note taking is correlated with an increase in the variance between initial and post-task usefulness judgements.

It is hypothesised that the extraneous cognitive load imposed affects the way heuristics are modified (*see* Section 7.2). After completing a task participants were asked to evaluate the usefulness of the notes that they made for each document. Notes made by participants can be seen as representative of their heuristic approach taken at a specific point in time. Heuristics change over time and the notes made by an individual may be invalidated or become less useful when an individual applies modified heuristics.

Table 8.5 compares task performance with note evaluation. Participants who demonstrated a basic or satisfactory understanding assigned an almost equal distribution of ticks, question marks, and crosses. This indicated that the extraneous cognitive load imposed on these individuals was high enough to prevent the extensive modification of heuristics used to address the task. 52% of participants who achieved a good understanding assigned crosses. This very high proportion of cross assignments indicated that it is likely that extensive modification of heuristics occurred. This may have been due to lower levels of extraneous cognitive load. Modification of heuristics is essential to enable information objects to be processed in a way that is compatible with an individual's evolving knowledge schemata.

MEAN AVG. PER TASK	BASIC OR SATISFACT. UNDERSTANDING	GOOD UNDERSTANDING
Tick (N=49)	(34/138)x100 = 24.638%	(15/102)x100 = 14.706%
Question Mark (N=45)	(35/138)x100 = 25.362%	(10/102)x100 = 9.804%
Cross (N=82)	(29/138)x100 = 21.014%	(53/105)x100 = 51.961%

TABLE 8.5: EVALUATION ANALYSIS

8.1.3 DISCUSSION

Exploratory Study 2 attempted to find evidence enabling the development of intrinsic (*imc*) and extraneous (*emc*) measures of comparison. These measures can be used to enable contributors to a model of cognitive load to be identified. As discussed in Section 7.2, both intrinsic and extraneous accounts of cognitive load need to be represented within a model.

Results show that a usefulness judgement can be regarded as a measure of how an individual interprets an information object at a specific point in time. Figures 8.1 and 8.2 show that initial and post-task usefulness judgements do not seem to be affected by successive evaluations and re-evaluations. Usefulness judgements did not significantly improve or worsen as more documents were browsed. Furthermore, Figure 8.3 and Table 8.2 show that initial and post-task judgements were assessed independently of each other. It is hypothesised that the intrinsic cognitive load imposed on an individual limits the type of heuristics deemed suitable by that individual for addressing their information need (*see* Section 7.2). The interpretation of information changes over time as an individuals heuristics change. When an individual applies heuristics that are *thought to be* more suited to resolving their information need, new interpretations of the usefulness of an information object are inevitable. Therefore, it can be concluded that the degree of variance between initial and post-task usefulness judgements is a suitable comparative measure of intrinsic cognitive load.

If an individual has well developed knowledge schemata, more intrinsic elements can integrated within an individual's working memory. Table 8.3 shows that participants who developed better knowledge schemata found it easier to assess the usefulness of documents after a search had been completed. This does not provide an insight into the cognitive load imposed when an initial relevancy judgement was made. Figure 8.4, however, shows that an increase in note taking is positively correlated with an increase in the variance between initial and post-task usefulness judgements. This demonstrates that intrinsic cognitive load is measurable.

It has also been found that the extraneous cognitive load imposed on an individual is identifiable. Notes made by participants can be seen as representative of their heuristic approach taken at a specific point in time. Heuristics change over time and the notes made by an individual may be invalidated or become less useful when an individual applies modified heuristics. If extensive modification of heuristics occurs, this increases the likelihood of notes being evaluated as being less useful after a task has been completed. Table 8.5 shows that during evaluation, participants who assigned a high proportion of 'cross assignments' performed better in tasks. However, this does not provide an insight into the extraneous cognitive load imposed when an initial judgement is made. A procedure enabling measurement is proposed in Section 9.2.

Exploratory Study 2 has shown variations in initial and post task usefulness judgements do seem to be indicative of the cognitive load imposed on an individual. Furthermore, both intrinsic and extraneous cognitive load accounts can be differentiated. These findings satisfy the aims of this study. However, to enable measurement of cognitive load, further comparative measures need to be developed.

The extent that cognitive load might have compromised the usefulness / usability of relevance feedback systems has not been explicitly investigated. This study did not require participants to use a relevance feedback system. The underlying process of feedback has not been the source of much debate or theory-building research in information science literature. This study encouraged participants to make notes during information seeking. The process of making notes can be seen as an individualistic feedback system. Examining how individuals generate feedback, for their own cognitive purposes, may reveal some of the underlying processes of feedback that cannot be studied when submitting feedback to an IR system.

8.2 EXPLORATORY STUDY 3: BEHAVIOURAL PATTERNS

Exploratory Study 3 aimed to investigate behavioural patterns associated with individuals who regularly submit passage feedback judgements. Evidence is needed that in genuine information seeking environments, cognitive load can be seen to affect behaviour. The way in which cognitive load affects feedback systems obviously needs to be observable. Before it is possible to model how cognitive load affects behaviour, the ability of individuals to self-report behavioural change has to be investigated. The postulate that behavioural variables can be identified by individuals as having an affect on making feedback judgements needs to be tested.

Aim:

- To demonstrate whether individuals can identify different behavioural patterns associated with a variety of information seeking scenarios.

Objectives needed to fulfil aim:

A. Discover if it is recognised how uncertainty, domain knowledge, and knowledge obtained from search interactions, affect relevance feedback.

B. Discover the most important (self-reported) factors that influence relevance feedback behaviour.

8.2.1 METHOD

21 individuals from a pool of 75 Kenjin[25] search engine users answered a questionnaire via email. The questionnaire was sent in December 2001 (*see* Appendix C). Kenjin is a commercially available retrieval tool that allows users to highlight passages within Web pages that they consider useful. These passages are then submitted to the Kenjin server. The Kenjin database contains user profiles that consist of submitted passages and the Web page addresses they were submitted from. When a user submits a passage, all user profiles are checked for similar passage submissions. Kenjin then assigns similarity scores between user profiles that contain similar passage submissions. The most popular Web sites visited by users that have been identified as having compatible interests (i.e., similar profiles) are returned. Users of Kenjin are likely to be expert in providing passage feedback. This was the reason behind the Kenjin user group being targeted.

[25] The developers of Kenjin provided e-mail addresses of 75 Kenjin users. These users were selected on the basis that they were heavy users and had previously expressed their interest in participating in user surveys.

8.2.2 RESULTS

Background Data

21 volunteers responded to the questionnaire: 16 males, and five females. Participants were from seven different occupational backgrounds. Six were students. All participants regularly used IR systems, mainly Web search engines. All of the participants claimed they were expert at providing an IR system with passage relevance feedback.

Uncertainty

Table 8.6 shows that 90% of the users surveyed believed that uncertainty always or frequently affected their highlighting feedback strategy. 71% suggested that as a result of uncertainty, they highlighted fewer passages. These findings provide evidence that uncertainty affects an individual's ability to provide feedback. Of the users who believed that uncertainty always or frequently affected their highlighting strategy, 87% suggested that uncertainty affected an individual's ability to submit highlighted passages for feedback.

1. If you are uncertain that highlighting a passage will provide good results does this affect your highlighting strategy?

Always affects my approach	11	52.38%
Frequently affects my approach	8	38.10%
Generally does affect my approach	2	9.52%
Sometimes affects my approach	0	0.00%
Generally does not affect my approach	0	0.00%
Rarely affects my approach	0	0.00%
Never affects my approach	0	0.00%
Don't know	0	0.00%
Total	21	100.00%
What do you do differently?		
Highlight more passages	3	14.29%
Highlight fewer passages	15	71.43%
Select your passages more carefully	1	4.76%
Other / Don't know	2	9.52%
Total	21	100.00%

TABLE 8.6: UNCERTAINTY

Table 8.7 indicates only 57% believed that domain knowledge always or frequently affected their highlighting feedback strategy. 58% suggested that a lack of domain knowledge encourages more passages to be highlighted. These findings do not suggest that an individual's domain knowledge has a strong effect on highlighting strategy. However, of the 57% who believed that domain knowledge always or frequently affected their highlighting strategy, 86% suggested that this encourages more passages to be highlighted. This tentatively suggests that some individuals may have developed a specific strategy for coping with a lack of domain knowledge.

2. If you are unfamiliar with the subject area you are searching does this affect your highlighting strategy?		
Always affects my approach	6	28.57%
Frequently affects my approach	6	28.57%
Generally does affect my approach	0	0.00%
Sometimes affects my approach	2	9.52%
Generally does not affect my approach	3	14.29%
Rarely affects my approach	3	14.29%
Never affects my approach	1	4.76%
Don't know	0	0.00%
Total	21	100.00%
What do you do differently?		
Highlight more passages	7	58.33%
Highlight fewer passages	3	25.00%
Select your passages more carefully	1	8.33%
Other / Don't know	1	8.33%
Total	12	100.00%

TABLE 8.7: UNFAMILIARITY

Results indicated the way in which previous unsuccessful search interactions affect highlighting strategy is unclear. Table 8.8 shows that responses were evenly distributed amongst categories. Previous search interactions are likely to cause the modification of heuristics used to address an information need. The modification of heuristics is likely to affect a wide range of behaviour rather than having a specific predictable effect.

3. If you have searched a subject area before and failed to find useful information does this affect your highlighting strategy when performing another search on the same subject area?

Always affects my approach	6	28.57%
Frequently affects my approach	4	19.05%
Generally does affect my approach	2	9.52%
Sometimes affects my approach	3	14.29%
Generally does not affect my approach	0	0.00%
Rarely affects my approach	3	14.29%
Never affects my approach	1	4.76%
Don't know	2	9.52%
Total	21	100.00%
What do you do differently?		
Highlight more passages	2	16.67%
Highlight fewer passages	1	8.83%
Select your passages more carefully	5	41.67%
Other / Don't know	4	33.33%
Total	12	100.00%

TABLE 8.8: PREVIOUS INTERACTIONS

Table 8.9 shows 87% of respondents believed that subject expertise always affected their highlighting strategy. Of this 87%, 79% believed that this enabled passages to be highlighted more carefully. These findings strongly support the idea that experts are able to provide more useful feedback. Possessing well-developed knowledge schemata enables greater feedback judgement accuracy at search time.

4. If you consider yourself to have good knowledge about a subject area you regularly search does this affect your highlighting strategy?

Always affects my approach	18	86.71%
Frequently affects my approach	0	0.00%
Generally does affect my approach	2	9.52%
Sometimes affects my approach	1	0.00%
Generally does not affect my approach	0	0.00%
Rarely affects my approach	0	0.00%
Never affects my approach	0	0.00%
Don't know	0	0.00%
Total	21	100.00%
What do you do differently?		
Highlight more passages	3	15.00%
Highlight fewer passages	1	5.00%
Select your passages more carefully	14	70.00%
Other / Don't know	2	10.00%
Total	20	100.00%

TABLE 8.9: EXPERTISE

Table 8.10 shows that 67% suggested that knowledge obtained from previous searches makes it easier to highlight passages. This finding is interesting as it implies that online search experience, rather than domain knowledge, is more important when making feedback judgements. A number of different reasons were given for why it is difficult to highlight passages. 29% suggested that they found it most difficult to highlight passages when 'there is too much information'. This response was the most frequent reason given and was free reported (not elicited). It seems that there may be a frequent occurrence of information seeking scenarios where cognitive load (or cognitive overload) makes it hard to provide feedback.

Part B

1. When do you find it easiest to highlight a passage?

Free responses were categorised:		
Knowledge from previous searches	14	66.67%
Other / Don't know	4	19.05%
When subject area is familiar	3	14.29%
Total	21	100.00%

2. When do you find it most difficult to highlight a passage?

Free responses were categorised:		
When there is too much information	6	28.58%
When browsing casually	5	23.81%
Uncertain that highlighting will help	4	19.05%
When subject area is complicated	4	19.05%
Other / Don't know	2	9.52%
Total	21	100.00%

TABLE 8.10: FREE RESPONSE

8.2.3 DISCUSSION

Before it is possible to model how cognitive load affects behaviour, the ability of individuals to self-report behavioural change has to be investigated. The postulate that behavioural variables can be identified by individuals as having an affect on making feedback judgements has been explored. Exploratory Study 3 showed that participants were able to self-report changes in relevance feedback behaviour. Participants recognised that different approaches are needed depending on the scenario imposed by an information need.

Although only simplistic behavioural changes were explored, it was discovered that individuals do adopt different approaches: uncertainty results in less passages being highlighted (*see* Table 8.6); low levels of domain knowledge often causes an increase in the number of passages highlighted (*see* Table 8.7); and expertise in a subject area allows passages to be selected more carefully (*see* Table 8.9).

A range of behavioural variables were identified by participants as having an affect on making feedback judgements (*see* Table 8.10). Categorisation of free responses helped to define six factors (*see* Table 8.11) that should be considered for inclusion within a model of cognitive load[26]. It would obviously be impossible to investigate all known behavioural variables identified within information science literature. Exploratory Study 3 has helped to identify behavioural variables that could *potentially* be incorporated within a cognitive load model of feedback. The factors listed in Table 8.11 have been the subject of extensive research within information science (*see* Chapters 2-4). It is not suggested that the six factors are the only ones that are suitable for potential inclusion within a model of cognitive load. However, it is hoped that they will provide a valuable insight. Investigating how these variables can be elicited from individuals and assessed for inclusion within a model of cognitive load, is detailed within Section 9.5.

---

[26] The development of intrinsic (*imc*) and extraneous (*emc*) measures of comparison will be used to assess whether these variables are suitable contributors to a model of cognitive load.

FACTORS TO BE ASSESSED	FREE REPORTED SCENARIOS THAT WERE IDENTIFIED AS CAUSING BEHAVIOURAL CHANGE
1. Knowledge obtained from interactions	Knowledge from previous searches
2. Information overload	When there is too much information
3. Exploratory v goal-directed behaviour	When browsing casually
4. Uncertainty	Uncertain that highlighting will help
5. Task complexity	When subject area is complicated
6. Domain knowledge	When subject area is familiar

TABLE 8.11: FACTORS FOR POTENTIAL INCLUSION WITHIN A MODEL OF COGNITIVE LOAD

8.3 OUTSTANDING ISSUES

- Initial and post-task usefulness judgements assigned to documents can be assessed on an individual basis, independently of other documents. No confounding variables were found that could be associated with the idea that initial and post-task judgements may improve or worsen as more documents are presented. Furthermore, it was found that an initial judgement does not significantly influence a post-task judgement.

- Poorly developed knowledge schemata cannot be fully integrated within working memory. Exploratory Study 2 set simulated task difficulty at a level that induced cognitive load. The majority of participants found that making notes helped them to manage tasks. Elements (separate ideas 'talked about') were included within notes because they could not be incorporated fully within existing knowledge schemata. Evidence has been found to demonstrate that cognitive load is a measurable phenomenon. The number of elements incorporated within notes was significantly positively correlated with the degree of variance between initial and post task judgements.

- A usefulness judgement can be seen as a measure of how an individual interprets an information object at a specific point in time. It can be suggested that the degree of variance between initial and post-task usefulness judgements is a suitable comparative measure of intrinsic cognitive load. However, the degree of variance needs to be correlated with other variables, to provide an insight into the intrinsic cognitive load imposed when an initial usefulness judgement was assigned. Exploratory Study 2 used the number of elements incorporated within notes as a correlative measure. However, to enable a fuller understanding, a range of behavioural variables and observable manifestations should also be correlated.

- When evaluating notes, participants who demonstrated a basic or satisfactory understanding assigned an almost equal distribution of ticks, question marks, and crosses. This indicates that the extraneous cognitive load imposed on these individuals was high enough to prevent the extensive modification of heuristics used to address the task. 52% of participants who achieved a good understanding assigned crosses. This very high proportion of cross assignments indicates that extensive modification of search heuristics was enabled due to a lower extraneous cognitive load imposed.

- Exploratory Study 3 found that participants were able to self-report changes in relevance feedback behaviour. Participants recognised that different approaches are needed depending on the scenario imposed by an information need. Although only simplistic behavioural changes were explored, it was discovered that individuals do adopt different approaches: Uncertainty results in less passages being highlighted; Low levels of domain knowledge often causes an increase in the number of passages highlighted; Expertise in a subject area allows passages to be selected more carefully.

- Exploratory Study 3 has helped to identify behavioural variables that could *potentially* be incorporated within a cognitive load model of feedback. These factors have been the subject of extensive research within information science: Knowledge obtained from interaction, information overload, exploratory v goal-directed behaviour, uncertainty, task complexity, and domain knowledge.

9.0 INTRODUCTION

This chapter outlines the investigative methods and strategies developed for the main experimental procedure used within this thesis (*see* Chapters 10 and 11). The goal of the main experimental procedure was to develop a model of cognitive load for relevance feedback. This model would be used to predict how the factors that have a major impact on levels of cognitive load affect the utility of feedback judgements.

The intention was not to set up a situation where participants were required to directly self-report how cognitive load affected their information behaviour. The nature of cognitive load makes self-report an invalid approach. Any self-report is burdened by cognitive load (*see* Section 7.1.1). Ellis (1996) believed that while it is feasible to analyse changes in knowledge in response to new information qualitatively, to attempt to do the same quantitatively seems to have no tenable theoretical foundation and represents an unsustainable research goal. The mechanisms developed by this thesis used to estimate the amount of cognitive load imposed on an individual, and to identify the factors causing cognitive load, do not attempt to measure changes in an individual's knowledge directly. This is another unsustainable research goal. Instead, the state of an individual's knowledge was ascertained by exploring factors that have been identified by qualitative research as being both measurable and able to provide an insight into an individual's information behaviour (*see* Chapter 4).

Ellis (1996) believed that the desire to simplify measurement as opposed to maintaining the complexity of phenomena has led to the pursuit of quantification at the expense of validity both inside and outside the experimental environment. The model proposed within this thesis cannot be used to provide detailed guidelines for the design of IR systems. However, establishing how human cognitive architecture (i.e., working memory) inhibits interaction with information objects shows that many observable complex behavioural patterns can be explained without the artificial simplification of phenomena.

Schamber (1994) identified that the studies performed by Barry (1994); Cool et al (1992); Schamber (1991); and Thomas (1993), were generally inspired by the idea that relevance judgements alone cannot convey the multiple meanings underlying users' decisions on whether to pursue information. This thesis was inspired by the idea that the utility of judgements maybe further inhibited by cognitive load. This thesis cannot be classified as being a cognitive study, a user centred study, a behavioural study, or a pragmatic / phenomenological study. The approach taken by this by thesis was exploratory.

Qualitative information derived from user explanation of their situation and decisions was used to perform a quantitative analysis that enables causal explanations. Although cognitive load cannot be self-reported, some variables such as uncertainty can be quantified (Wilson et al, 2000). A causality analysis requires a model in order to function. The causality analysis performed within this thesis **does not** involve testing a pre-defined cognitive model of a user that attempts to predict that effect information has on a user's knowledge state. Ellis (1984) suggested that it is insufficient and unsatisfactory to argue that because there are hosts of uncontrollable variables for a researcher, these cannot be taken into account during laboratory testing. Ellis believed that this is a way of transforming a theoretical problem concerning explanatory power into an alleged methodological problem associated with uncontrollable variables. The aim of the model developed within this thesis was to explain how a user's behavioural patterns are affected by cognitive load. This was achieved by evaluating the contribution uncontrollable variables have on cognitive load.

9.1 INTRINSIC MEASURES OF COMPARISON

The exploratory studies detailed above (*see* Chapter 8) provided enough evidence to allow the development of an intrinsic[27] (*imc*) measure of comparison. This measure can be used to enable contributors to a model of cognitive load to be identified.

[27] *See* Sections 7.2 and 8.1.3 for explanations of intrinsic cognitive load.

Equation 9.1 (on page 145) was used to estimate the degree of intrinsic cognitive load during the search process. The degree of intrinsic load is the variance between usefulness judgements assigned to a document (by a searcher) during and after the search process. A relevant document is not necessarily pertinent to a user's information needs (*see* Section 2.1). During exploratory experimentation, participants found that evaluating a document using a usefulness magnitude scale was straightforward (*see* Section 8.1.1). Assigning a measure of usefulness is analogous to evaluating the pertinence of a document.

Intrinsic cognitive load is considered as a barrier to providing an accurate usefulness judgement during the search process. When an initial usefulness judgement is made for document j, and compared to a post-search usefulness judgement for document j, intrinsic load is calculated. The assumption that a post-search usefulness judgement is more accurate than a usefulness judgement made during the search process, due to cognitive load, has been tested by Exploratory Study 2 (*see* Section 8.1.3). Finding explanations for the variance between judgements can be achieved by exploring the factors that can potentially instigate behavioural change (*see* Section 9.5).

During experimentation, the intrinsic cognitive load imposed on an individual can be calculated for each individual document (information object). A comparative analysis of intrinsic cognitive load measures across a range of tasks/participants cannot be performed without normalisation. As previously reported in Section 8.1.3 participants who only achieved a satisfactory or basic understanding had more residual cognitive load, so found it harder to identify which documents were useful when a task has been completed. Equation 9.1 incorporates a baseline task performance normalisation. A high level of residual intrinsic cognitive load is inevitable if task performance is poor. A low level of residual intrinsic load is expected when task performance is good.

Normalisation procedure:

- If task performance is good, then baseline ≈ 1. No normalisation is required as the individual is likely to have developed good knowledge schemata. Sophisticated knowledge schemata free working memory and result in a low residual intrinsic cognitive load after a task has been completed. The ability to perform post-task evaluations is therefore largely uninhibited.

- If a task is not successfully resolved but a participant is able to demonstrate a satisfactory or basic understanding of the problem space, then baseline > 1. Normalisation is required as the development of sophisticated knowledge schemata is less likely. The ability to perform post-task evaluation is therefore inhibited by residual intrinsic cognitive load. It is still possible to evaluate the load imposed during initial usefulness judgements as normalisation accounts for the amount of residual intrinsic load. Evidence of a satisfactory or basic understanding of the problem space means that the residual intrinsic cognitive load is likely to be lower than the load imposed during information seeking.

- If a participant is unable to demonstrate a basic understanding of the problem space, an evaluation of the load imposed during initial usefulness judgements is not possible. The significant development of knowledge schemata is unlikely. Therefore, residual intrinsic cognitive load ≈ cognitive load experienced during information seeking.

$$imc_j = \frac{icl}{bicl} = \frac{\left|iuj_j - puj_j\right|}{\left(\dfrac{puj_j}{tfe}\right)}$$

Where

icl = Intrinsic Cognitive Load

$bicl$ = Baseline Intrinsic Cognitive Load

iuj_j = Initial Usefulness Judgement assigned to document j

puj_j = Post-task Usefulness Judgement assigned to document j

tfe = Task Fulfilment Evaluation

EQUATION 9.1

How the task fulfilment evaluation (*tfe*) measure was derived is described in Section 9.3.

9.2 GENEPLORE MODEL AND FEEDBACK PROCEDURE

The extraneous[28] cognitive load imposed on an individual affects the way in which heuristics are modified. Modification of heuristics is essential to enable information objects to be processed in a way that is compatible with an individual's existing knowledge schemas. A high level of extraneous cognitive load prevents the extensive modification of heuristics. Examining the modification of an individual's heuristics can be achieved by encouraging experiment participants to make notes during information seeking.

Poorly developed knowledge schemata cannot be fully integrated within working memory. Individuals make notes if they cannot incorporate elements within working memory. Notes made by individuals are representative of their heuristic approach taken at a specific point in time. Heuristics change over time and the notes made by an individual may become invalid or less useful when an individual applies modified heuristics.

[28] *See* Sections 7.2 and 8.1.3 for explanations of extraneous cognitive load.

During the second stage of experimentation, reported in Section 8.1.3, experiment participants were asked to evaluate the usefulness of the notes that they made for each document. Participants who demonstrated a basic or satisfactory understanding assigned an almost equal distribution of ticks, question marks, and crosses. This tentatively indicated that the extraneous cognitive load imposed on these individuals was high enough to prevent the extensive modification of heuristics used to address the task. Participants who achieved a good understanding assigned a significantly higher proportion of cross assignments (sig. at the 0.01 level). It can be suggested that a greater number of cross assignments showed that extensive modification of search heuristics was enabled due to a lower extraneous cognitive load imposed.

The Exploratory Study 2, reported in Section 8.1, used a procedure for evaluating extraneous cognitive load that needs to be improved upon. There are two significant problems associated with the procedure. Firstly, the validity of relying on participants to make notes that accurately reflect the number of elements that they cannot incorporate within working memory is questionable. Secondly, the procedure did not explicitly explore the degree to which extraneous cognitive load affected the generation of feedback judgements. Instead the procedure was an attempt at evaluating extraneous cognitive load as a barrier to task performance, the difficultly associated with the generation of feedback judgements being treated as implicit.

When investigating extraneous cognitive load, a mechanism is needed that allows changes in an individual's heuristics during the process of information seeking to be discovered within the context of providing feedback judgements. It is proposed that the required mechanism should allow participants to construct a written answer to a simulated retrieval task. A participant's self-reported evaluation of the answer construction process enabled extraneous load to be explored. The answer construction process enabled conceptual elements from passages submitted for feedback purposes to be incorporated by participants. Issues associated with enforcing participants to make notes were avoided. Participants had to provide a written answer or they failed to complete the task. Participants were not informed of the number of documents that were to be presented; this encouraged them to attempt an answer after browsing each document. Each attempt at an answer was considered representative of the heuristics they applied at a specific point in time.

Developing an answer to a simulated task is a creative process. Changes to an individual's heuristics are inherent within the creative process. It is not obvious how to study creativity under controlled laboratory conditions. The subject of creativity has had unscientific connotations, perhaps resulting from the reliance on introspective anecdotal accounts in previous attempts to describe the creative process (e.g., Ghiselin, 1952).

Establishing general cognitive principles of creativity that apply across many domains is difficult. However, there is no reason why the patterns that enable the analysis of creativity in one domain cannot be extended to other domains as long as the strategies are sufficiently flexible to accommodate changes (Finke, Smith, and Ward, 1996)[29]. An important part of being creative is knowing 'the rules of the game' and becoming skilled at identifying them (Perkins, 1981).

The experimental procedure used to investigate extraneous load adhered to the following:

> It is important, when exploring cognition, to structure tasks in a way that subjects have the opportunity to make genuine discoveries within experimental context; this can be done without compromising the scientific integrity of the research. In addition, it is much easier to predict when something will be creative than to predict exactly what form the creation will take... Creative performance [should not] be explained simply in terms of 'creative people do creative things because they think in creative ways'... It [should be] explained with reference to particular kinds of structures that a person employs and in terms of the properties of these structures... The most reasonable approach is to seek out general principles of creative thought, expressed in terms of behavioural processes and structures as opposed to trying to reduce creative cognitions to the level of computational units and algorithms (Finke, Smith and Ward, 1996, p 6-7).

[29] Having a general theory of creative cognition does not mean that it is possible to predict when a creative idea will occur.

Information scientists and psychologists have researched aspects of cognitive style (e.g., Ford, 2000), convergent and divergent thinking (e.g. Kelly, 1963), accepting change and tolerating ambiguity (e.g., Kuhlthau, 1993a; 1993b). These important behavioural variables, acting simultaneously, induce the complexities of information behaviour. Psychologists who have studied creativity have explored the same set of behavioural variables. Guilford (1956; 1968) was one of the first to study the structure of intellect by adopting a psychometric approach. He developed various tests to distinguish between convergent and divergent thinking. There are a variety of cognitive styles thought to be related to creativity, for example: Shouksmith (1970) investigated abstract modes of thought, Amabile (1983) explored the use of wide rather than narrow conceptual categories, and McLeod and Cropley (1989) found that the ability to detect discrepancies, discovering analogies, overcoming habitual patterns, accepting novelty, and tolerating ambiguities, all contribute to creativity.

To enable the study of extraneous load imposed on an individual, a procedure is needed that can model creative processes while still allowing sufficient flexibility for creative ideas to occur. The Geneplore model developed by Finke, Smith, and Ward (1996) considers both generative and exploratory cognitive processes. It provides an excellent framework that can be used to describe the basic cognitive processes related to creativity. In the initial, generative stage, mental representations called preinventive structures are constructed. The properties of preinventive structures are then exploited during an exploratory phase. If initial explorations result in a satisfactory resolution to the task at hand, the initial preinventive structure may lead directly to a creative product. If explorations are unsuccessful, one of two procedures would come into play, either of which would involve a return to the generative phase: abandon the initial preinventive structures and generate another that may be more promising, or modify the initial structure and then repeat the exploratory phase with the modified structure. By continuing these procedures, it is possible to gradually focus on the emergent structure of particular themes or problems or expand the structure to explore more general conceptual possibilities. The Geneplore model also considers constraints that can be imposed during either the generative or explorative stage. Some individuals may be more skilled at generating preinventive structures and others may be more skilled at interpreting them. This may help to explain why there are often dramatic individual differences in cognitive style.

Table 9.1 includes important (verifiable) cognitive processes, structures, properties and constraints, as defined by the Geneplore model. It includes a novel feedback procedure that is used to investigate extraneous load. This feedback procedure was developed to be compatible with the Geneplore model.

Generative Processes	**Pre-inventive Structures**	**Pre-Inventive Properties**	**Exploratory Processes**	**Feedback Procedure**
Retrieval	N/A	Novelty	Attribute Finding	Passage Selection
Association	N/A	Ambiguity	Conceptual Interpret-ation	Contextual Understanding
Synthesis	Mental Blends	Meaning-fulness	Functional inference	Utility Evaluation
Trans-Formation	Category Exemplars	Emergence	Contextual Shifting	Passage Trans-figuration
Analogical Transfer	Mental Models	Incongruity	Hypothesis Testing	Answer Construction
Categorical Reduction	Verbal Combination	Divergence	Searching for Limitations	Answer Evaluation

TABLE 9.1: THE GENEPLORE MODEL AND FEEDBACK PROCEDURE

Generative Processes within the Geneplore Model

The most basic type of generative processes consist of the *retrieval* of existing structures from memory and the formation of *associations* among these structures. Typically, these retrieval and associative processes happen quickly and automatically, but sometimes they are inhibited, i.e., by cognitive load. A rich variety of preinventive structures results from the mental *synthesis* of component parts and by the mental *transformation* of the resulting forms (schemata). Single concepts can be combined to form more complex concepts, with the meanings of initial concepts being altered as a result. Another type of generative process is *analogical transfer* in which a relationship in one context is transformed to another resulting in structures that are analogous to those already familiar. *Categorical reduction* is the process of mentally reducing objects or elements to more primitive categorical descriptions.

The generative processes shown in Table 9.1 have been explored within cognitive psychology. McClelland and Rumelhart (1986) found that parallel processing is more likely to be predominate when the preinventive structures are initially assembled. This suggests that the affect of cognitive load on an individual maybe greater during synthesis, transformation, analogical transfer, and categorical reduction, rather than during retrieval and association. Schneider and Shiffrin (1977) provide further evidence by arguing that as preinventive structures become more developed (i.e., during schema induction), processes become more focussed and controlled.

Preinventive Structures within the Geneplore Model

A *mental blend* refers to a class of structures that include conceptual combinations, metaphors, and blended mental images. What they have in common is that two or more distinct entities have been fused to create something new. *Category exemplars* consist of hypothetical categories that have features that are in common with familiar categories along with novel or emergent features that lead to new unanticipated discoveries. *Mental models* usually start as structures that are incomplete or unstable and are then improved and refined with exploration and discovery (Gentner and Stevens, 1983). Lastly, preinventive structures can consists of various types of *verbal combinations* that differ from mental blends in that the elements need not actually be fused physically in the structure.

Preinventive Properties within the Geneplore Model

Although a familiar structure might be interpreted in creative ways, the possibilities for creative discovery should be much greater if the structure exhibits *novelty.* Wisniewski and Gentner (1991) found that *ambiguity* contributes to discoveries that are made in many types of conceptual combinations. Preinventive structures often seem to have an underlying *meaningfulness* to them, which encourages further exploration and search. *Emergence* refers to the extent unexpected features appear in the preinventive structure. *Incongruity* refers to conflict among elements, encouraging further exploration to uncover deeper meanings (Wisniewski and Gentner, 1991). *Divergence* is related to ambiguity but refers specifically to the capacity to find multiple uses for the same structure.

Exploratory Processes within the Geneplore Model

Attribute finding is the systematic search for emergent features or can be used for exploring emergent features from the creation of conceptual combinations and metaphors. *Conceptual interpretation* is the process of finding abstract, metaphorical, or theoretical interpretations. More generally, conceptual interpretation can be thought of as the application of world knowledge to the task of creative exploration. *Functional inference* is facilitated by the exploration of potential uses. *Contextual shifting* is a way of considering a structure in new or different contexts as a way of gaining insights to possible uses or meanings. *Hypothesis testing* is when a structure is interpreted as being a possible solution to a problem. *Searching for limitations* can help to restrict future explorations and explore the implications of the problem.

Feedback Procedure

The pragmatics of the feedback procedure is discussed in Section 9.3. A theoretical justification of how the feedback procedure is compatible with the Geneplore model is detailed below. Words in italics refer to previously discussed features of the Geneplore model.

1. The process of *retrieving* existing structures from memory is analogous to selecting a passage from a document for feedback purposes. It can be assumed that the reason why a participant selects a passage is because it has a *novel* property, which in turn facilitates *the finding of attributes*, which may enable the creation of conceptual combinations that are needed for ***associations*** to be made (*see* below).

2. The process of ***association*** can be investigated by asking a participant if they fully understand the contextual meaning of the selected passage. It can be assumed that if a participant suggests that they do not fully understand, then this reveals *ambiguity*. *Conceptual interpretation* is likely to be triggered by ambiguity, which in turn may enable ***synthesis*** (*see* below).

3. The process of ***synthesis*** can be investigated by asking a participant to evaluate the utility (in terms of addressing their information need) of a selected passage. If a participant is able to perform an evaluation of utility, this indicates that a *mental blend* had occurred (as the evaluation of a specific passage is performed with respect to all other passages and existing knowledge). If a passage is deemed only to be potentially useful, then this indicates that a search for *meaningfulness* is underway. *Functional inference* is likely to be triggered by a search for meaningfulness, which in turn may enable ***transformation*** (*see* below).

4. The process of ***transformation*** is represented by the way a participant transfigures a selected passage, i.e., modifies the semantics by altering the feedback passage text when constructing an answer to the task. This indicates the existence of *category exemplars*. The process of *emergence* allows for *contextual shifting*, which in turn may result in ***analogical transfer*** (*see* below).

5. The process of ***analogical transfer*** can be investigated by asking a participant to construct an answer using modified passages (i.e., writing a solution to the task). This indicates the existence of *a mental model.* An *incongruity* within the mental model is likely to *trigger hypothesis testing*, which allows for ***categorical reduction*** (*see* below).

6. The process of ***categorical reduction*** can be investigated by asking a participant to evaluate their solution to the information problem. This evaluation process requires the need for the analysis of *verbal combinations.* A *divergence* between mental and verbal models is likely to trigger *a search for limitations.*

An Evaluation of the Geneplore Model

The Geneplore model provides an attractive alternative to those based exclusively on general descriptions of cognitive influences and styles (Guilford, 1956; 1968). The model allows specific types of cognitive processes to be explored but need not be reduced to the level of strictly computational processes. The model allows individuals to discover concepts that turn out to be much more important than the ones they were contemplating at the beginning. This makes the model especially suited to the analysis of information behaviour, unlike many models that specify design parameters and also determine what form will satisfy those requirements. This flexibility is essential to the problem solving process (Bransford and Stein, 1984).

9.3 EXTRANEOUS MEASURE OF COMPARISON

Now that a feedback procedure has been specified (pragmatics outlined in Chapter 9), it is possible to develop the equation required for an extraneous measure of comparison (*emc*). This measure can be used to enable contributors to a model of extraneous cognitive load to be identified.

Equation 9.2 can be used to estimate the degree of extraneous cognitive load during the search process. The degree of extraneous load is the variance between the evaluation of usefulness of a feedback passage while browsing a document and during answer construction. Extraneous cognitive load is considered as a barrier to providing accurate feedback judgements during the search process. When an initial usefulness evaluation (*ie*) is made for passage *pj*, and compared to a usefulness evaluation (*pe*) for passage *pj* during answer construction, extraneous load is calculated. Finding explanations for the variance between judgements will be achieved by exploring the factors that instigate behavioural change.

Within the main experimental procedure providing feedback involves the use of an interactive highlighting tool. If a searcher judges a document passage as being useful, they are encouraged to submit it for feedback purposes. The utility of these passages can then be evaluated by considering the extent that they were used to construct an answer to a retrieval task. A brief description of how this evaluation process takes place is described below.

$$ecl = \left| ie_{pj} - pe_{pj} \right|$$

Where

ecl = Extraneous Cognitive Load

ie_{pj} = Initial evaluation of passage usefulness for *pj*

pe_{pj} = Evaluation of passage usefulness for *pj* during answer construction

EQUATION 9.2

Answer construction

When a participant submitted a passage for feedback purposes, he or she was asked to incorporate the passage within their answer to the simulated task[30]. Answer construction is a dynamic process. As the information seeking process progressed, participants were able to continually update and re-examine their answer. The number of times feedback passages could be removed, modified, and re-incorporated during answer construction was unlimited. Participants were encouraged to provide a concise answer using as many feedback passages as possible. Although adding complementary words / sentences was expected, it was advised that it would be preferable to attempt to construct an answer that made as much use of feedback passages as possible.

Initial evaluation of usefulness (*ie*)

- *ie* = 1

  On submission, participants were asked to identify if a feedback passages was already conceptually similar to the existing answer construction and therefore not suitable for incorporation. Submitted passages of this nature were assigned an *ie* score of 1. An *ie* = 1 score indicated that the submitted passage was, initially, not useful for answer construction.

- *ie* = 2

  On submission, participants were instructed to identify submitted passages that, at that specific point in time, could not be incorporated within their answer construction. Passages of this nature were assigned an *ie* score of 2. An *ie* = 2 score indicated that the submitted passage was initially considered to be potentially useful for answer construction at a later time (possibly when a participant's information need was more precisely defined).

[30] This was achieved using 'copy and paste' word processing tools.

- *ie* = 3

  If participants used a submitted passage for the answer construction process immediately, an *ie* = 3 score was assigned. An *ie* = 3 score indicated that this passage was initially very useful.

Evaluation of passage usefulness during answer construction (*pe*)

- *pe* = 0

  After an answer was constructed, if a passage was not used for the answer construction process, and **was not** identified as being conceptually similar to an element of the answer construction, a *pe* = 0 score was assigned.

- *pe* = 1

  After an answer was constructed, if a passage was not used for the answer construction process, and **was** identified as being conceptually similar to an element of the answer construction, a *pe* = 1 score was assigned. Passages of this nature were useful for feedback purposes as they could enable similar potentially more useful information objects to be retrieved.

- *pe* = 2

  If participants used a submitted passage for the answer construction process, but then removed or modified the semantics of the passage a *pe* = 2 score was assigned. A *pe* = 2 score indicated that passages of this nature significantly influenced the development of a participant's information need.

- *pe* = 3

  If participants used a submitted passage for the answer construction process, and did not extensively modify the semantics of the passage a *pe* = 3 score was assigned. A *pe* = 3 score indicated that passages of this nature were critical in addressing an aspect of a participant's information need.

9.4 MULTIPLE REGRESSION ANALYSIS

A range of behavioural factors can *potentially* affect the generation of feedback judgements by inducing cognitive load. As many factors as *possible* should be considered as potential contributors to a model of cognitive load (*see* Section 9.5). Now that the intrinsic (*imc*) and extraneous (*emc*) measures of comparison can be developed (*see* Sections 9.1 and 9.3), this enables prospective 'cognitive load model' contributors to be evaluated using a Stepwise multiple regression analysis[31].

Multiple regression is the process of constructing a linear equation that will predict the values of a target (dependent) variable from knowledge of specified values of two or more predictor (independent) variables. The objective of multiple regression is to determine if the addition of more predictor (independent) variables improves the accuracy of the prediction of the target (dependent variable). However, the issue of examining predictor variables and determining if some are more useful than others is problematic (Darlington, 1968). The basic problem with evaluating the associated importance of predictor variables is that correlation does not imply causation. Furthermore, if everything correlates with everything else, attributing variance in the dependent variable to any one independent variable is impossible.

In a multiple regression equation, the coefficients of the independent variables are known as partial regression coefficients. They express an increase in the dependent variable that would be produced by a positive increase of one unit in the independent variable concerned, the effects of the other independent variables, both on the independent variable and the dependent variable being supposedly held constant. Statistical control is, however, no substitute for experimental control, where the independent variable having been manipulated by the experimenter, really is independent of the dependent variable.

[31] In simultaneous multiple regression, all available independent variables are entered in the equation directly. In stepwise multiple regression, the independent variables are added or removed from the equation one at a time, the order of entry or removal being determined by statistical considerations. In forward selection, predictors are added one at a time provided that they meet an entry criterion and cannot subsequently be removed. In backward selection, predictors are all present initially and are removed one at a time if they do not meet a retention criterion. Stepwise is a combination of these processes.

If scores on all the variables in a multiple regression equation are standardised, the intercept of the regression equation disappears and each regression coefficient, referred to as a beta weight, expresses the change in the dependent variable, expressed in standard deviation units, that would be produced by a positive increment of one standard deviation in the dependent variable concerned. The addition of new predictors can radically affect the relative contribution of those variables already in the equation. When planning a multiple regression the selection of predictors must be guided by a substantive theoretical rationale. A statistical model cannot itself yield an unequivocal interpretation of regression results (Clark-Carter, 1998). A substantive model of causation is required. This model of causation is described in Section 9.5. The model of causation was based on recognised factors, identified by qualitative research as being both measurable and able to provide an insight into an individual's information behaviour.

The intrinsic (*imc*) and extraneous (*emc*) multiple regression analyses are shown below:

$$\text{Predicted } imc = b_1 X + b_2 Y + \ldots + b_0$$

Where

b_n = testing regression coefficients determines whether potential contributors to the model have a non-zero relationship to the intrinsic measure of comparison, i.e., this is a measure of the steepness of the best-fit lines.

b_0 = regression constant, i.e., the intercept.

EQUATION 9.3

$$\text{Predicted } emc = b_1X + b_2Y + ... + b_0$$

Where

b_n = testing regression coefficients determines whether potential contributors to the model have a non-zero relationship to the extraneous measure of comparison, i.e., this is a measure of the steepness of the best-fit lines.

EQUATION 9.4

Use of statistics software enabled a stepwise regression to be performed. Variables were placed in the model one at a time. After a new variable was added, the contribution of each variable in the model was reassessed. If it was determined that a variable present in the model did not contribute significantly, then it was removed. The F-ratio was used to determine if a variable could enter or be removed from the model (*see* Equation 9.5). The values of *F-to-enter* and *F-to-remove* were set at 0.1 and 0.5 respectively. The aim was to identify a small combination of IVs that account for the majority of variance in the DV.

$$F = \frac{\text{variance in DV explained by IV}}{\text{variance in DV not explained by IV}}$$

EQUATION 9.5

Intrinsic and extraneous measures of comparison avoid problems associated with the testing of pre-defined cognitive models, which attempt to artificially predict the affect information has on a user's knowledge state. The use of intrinsic and extraneous measures of comparison enabled qualitative information, derived from users' explanations of their situation and decisions, to lead to causal, quantitative explanations.

9.5 BEHAVIOURAL VARIABLES

Now that a model of cognitive load can be mathematically developed, the behavioural factors used to populate the model need to be stipulated. Six key factors were chosen: Knowledge obtained from search sessions, coping with information overload, exploratory v goal-directed behaviour, uncertainty, domain knowledge, and task complexity. These factors have been previously discussed in Chapter 4. Exploratory Study 3 has also helped to identify these factors (*see* Section 8.2.3). It is not suggested that the six factors are the only ones that are suitable for potential inclusion within a model of cognitive load. However, it was hoped that they would provide a useful insight/starting-point. Crucially, these factors needed to be elicited from experiment participants in a way that enabled them to be assessed during information seeking. Previous research has shown that the six factors chosen can be measured/assessed.

9.5.1 KNOWLEDGE OBTAINED FROM SEARCH SESSIONS

As previously discussed (*see* Section 3.4), the only way of estimating the impact information has on an individual's knowledge is to use a surrogate measure. Within this thesis, uncertainty was used as the surrogate measure for both conceptual and technical knowledge. Wilson (1999) suggested that the concept of uncertainty could be a useful analytical variable. However, the concept of uncertainty and how it relates to information behaviour is not well researched. The study conducted by Wilson et al (2000) concluded that the concept of uncertainty was recognised by users of IR systems and could be measured. Kuhlthau (1993a) suggested that during information seeking, there are two types of uncertainty: one of a conceptual nature (i.e., what the user is looking for) and one related to the interactive technical process of retrieving information[32].

[32] Technical knowledge should not be confused with a user's IR system expertise. Selecting an appropriate search strategy (or set of problem solving heuristics), suitable for satisfying an information need, is technical knowledge.

Wilson et al (2000) performed experimentation that revealed a further delineation. Uncertainty associated with 'affective dimensions', and cognitive uncertainty associated with an individual's perception of how a search is progressing (cognitive state uncertainty). Collating results from a large number of users reveal that individuals commonly experience a series of phases or stages as they seek information. However, this only provides an abstract view of the search process. It does not increase our understanding of how an individual progresses from one search stage to another. An evaluation of each individual's epistemic processes is needed to *understand* an individual's experience.

When examining an individual's epistemic processes, determining how much uncertainty is conceptual, and how much uncertainty is technical is difficult. Knowledge obtained from search sessions can be both conceptual and technical. When applying models of human information processing capabilities it is clear that delineating conceptual knowledge from technical knowledge is problematic. For example, the ACT* (Anderson, 1983) principle works on the basis that all knowledge is generated from declarative information. Declarative memory (conceptual knowledge) takes the form of a semantic net that attempts to link elements of information. Procedural memory (technical knowledge) attempts to assign a set of conditions/rules to each element of information, generating a production. Procedural knowledge (technical knowledge) is acquired by making inferences from existing declarative information (conceptual knowledge), thus making the delineation of conceptual knowledge from technical knowledge difficult.

The representation of knowledge obtained from search interactions within a model of relevance feedback is crucial. Interactive IR systems attempt to support a user's changing state of knowledge. In other words, they consider how a user's state of knowledge can influence the search process. The assumptions underlying the best-match paradigm mean that such systems cannot use information from the user about uncertainty or suspicion of inadequacy in the user's state of knowledge (*see* Chapter 2). The development of a relevance feedback system that supports a user's changing state of knowledge would need a model of situational usefulness that can be inferred from the user. Deriving such a model may be possible if the integral relationship between a user's relevance judgements and their movement through the information seeking process is examined.

A novel approach was needed to measure knowledge obtained from search interactions. Questions need to be posed to an individual after interaction with an information object. Questions must be easily understood within the context of information seeking. The following questions may enable the situational usefulness of an interaction to be evaluated:

1) How close do you think you are to completing the retrieval task?

Definition ------------------------------>------------------------------- Completion

2) Do you think that you will be able to complete the retrieval task?

Unlikely ------------------------------>------------------------------- Definitely

If a participant were asked to quantify their responses to the above questions on a magnitude scale, this would enable an individual's interpretation of the progress of a search to be continually updated.

- It was hypothesised that Question 1 would allow 'conceptual' and 'cognitive state' uncertainty to be measured. Question 1 enables an individual to self-report the sophistication of the knowledge schemata that they posses. It is hypothesised that well developed knowledge schemata enable both 'conceptual' and 'cognitive state' uncertainty to be lowered.

- It was hypothesised that Question 2 would allow 'technical' and 'affective' uncertainty to be measured. Question 2 allows an individual to self-report the appropriateness of the heuristics that they are currently applying.

9.5.2 COPING WITH INFORMATION OVERLOAD

A method for evaluating relevance judgements assigned to information objects must allow judges to signal their cognitive capability during evaluation (*see* Chapter 3). Relevance judgements alone cannot convey the multiple reasons underlying users' decisions on whether to pursue information. For a relevance judgement to be useful, it must accurately reflect a user's knowledge state. The concept of 'judgement accuracy' has not been researched by information scientists. If a user is struggling to cope with information overload, then the likelihood of accurate relevance judgements being made is lowered.

This thesis will avoid making the assumption that uncertainty is a suitable measure of an individual's ability to process information. The possibility that manifestations of uncertainty are symptomatic of limitations to cognitive capabilities needs to be considered. Researchers who have investigated how individuals cope with information overload have identified that the redundancy and uniqueness of information is an important aspect (e.g., Kelly, 1963; Huber and Roth, 1999). When evaluating how individuals cope with information overload, it is important to discover if a user's mood (invitational or indicative) results in a modification of information seeking behaviour (*see* Section 9.5.5).

A user's self-report of their invitational or indicative mood may also be associated with other affective dimensions such as motivation. A way of revealing causal factors exclusively associated with information overload is needed. The only obvious way of studying information overload is to develop a mechanism for measuring cognitive load. This mechanism may help explain why an individual has adopted a particular behaviour.

9.5.3 EXPLORATORY V GOAL-DIRECTED BEHAVIOUR

Borlund and Ingwersen (1997) developed a method of evaluation based on relevance assessments, which attempts to take into account the dynamic nature of information needs. Needs develop over time for an individual. The method is based on the introduction of a 'simulated work task situation' concept. Relevance assessments can be made with reference to concepts of situational as well as topical relevance. An individual's information expectations should be taken into account during the evaluation of an IR system. Borland and Ingwersen's method of evaluation aims to incorporate an individual's learning process, i.e., the individual's information need is allowed to shift focus during the process. The simulated work task situation acts as the point of reference against which situational relevance is measured (i.e., the relationship between the information object and the user's underlying need). An example of simulated information need situation from Borlund and Ingwersen (1997) is shown below:

Indicative request: Find something about Critical Success Factors (CSF)

Definition: Critical Success factors are, for any business, the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organisation.

Simulated work task situation: The boss in the import/export company in which you work, has told you to prepare a report which describes methods which can improve the company's competitiveness.

The simulated information need situation is characterised by an indicative request, then a definition of the request, and finally a simulated work task situation which describes the source of the need, the environmental situation and the problem which has to be solved. The objective is to ensure that relevance judgements are thus not based on the request or query put to the system, but rather relate to the user's perception of simulated task requirements and mental state during information seeking.

Ingwersen and Borland found that there was no significant differences between the way users modify a query when presenting their own needs compared to users who were given simulated needs. Within the experimental context applied by Ingwersen and Borland, it can be suggested that the simulated work task provided freedom for each individual to react in relation to their own interpretation of the information need. The main experimental procedure used within this thesis was conducted using simulated tasks that were formulated in the same way as Ingwersen and Borland's tasks (i.e., they consisted of a indicative request, definition, and work task situation). An individual's interpretation of an information need is likely to require a combination of exploratory and goal-directed behaviour. Simulated work tasks encourage both these types of behaviour. It would be interesting to discover how exploratory or goal-directed behaviour affects cognitive load. It might be found that adopting an exploratory or goal-directed behaviour enables an individual to manage the burden of cognitive load more effectively. Implementing system features that enable both exploratory and goal-directed behaviours to be reflected, allows greater flexibility of interaction rather than trying to predict what features should be available to a user depending on their cognitive state (*see* Chapter 4). A user may well be unable to communicate their cognitive state to a system. A user's perceived mental model may not be an accurate reflection of their information need situation. Offering flexibility of interaction may allow a user to realise that their perceived mental model may be inaccurate. A further justification of using simulated tasks was to ensure that cognitive load was induced by presenting complex information, which required multiple sources of information to be assimilated (*see* Section 10.3)

9.5.4 UNCERTAINTY

Wilson et al (2000) explored the suitability of uncertainty as an operational variable. No difference in the overall level of uncertainty expressed by different sexes or age groups were found. They discovered that experiment participants from different disciplines (academic departments) were able to identify with the problem solving stages proposed by Kuhlthau (1993b). No significant differences between disciplines existed. This finding enables comparisons to be made between levels of uncertainty experienced at different problem solving stages. However, Wilson et al could not perform this comparison due to the skewed nature of their distribution.

Surprisingly, Wilson et al found no significant difference in the level of uncertainty expressed by experiment participants before and after the online search. However, the data suggested that further exploration is necessary. Levels of uncertainty increased immediately following a search and then fell again at the follow-up stage. Participants were asked to state their level of uncertainty at three stages: pre-search, post-search, follow-up (Two months after). It would be interesting to explore fluctuations in uncertainty during IR interaction. The process of planning is the fundamental cognitive process that influences information processing (*see* Chapter 5). If dissonance occurs during IR interaction, then it is unlikely that information seeking behaviour will change unless the strength of the dissonance increases a user's level of uncertainty to a level that affects the cognitive load imposed on working memory. Exploring the extent fluctuations in uncertainty during information seeking affects cognitive load needs to be determined. This was an aim of the experimentation reported in Chapters 10 and 11.

9.5.5 DOMAIN KNOWLEDGE

Information elements exist in declarative and procedural memory and are not fully integrated within working memory. This crucial limitation to the processing of information should be considered when analysing information behaviour (*see* Chapter 6). The interpretation of perceptual clues, to determine if an information object is pertinent, is dependent on structures of goals held in memory and expertise. Importantly, identifying the characteristics of perceptual clues, where the discovery of novelty is an essential element, cannot be achieved from post-hoc generalisations. Experimentation needs to explore the reasons why cognitive load may prevent the identification of perceptual clues for relevance feedback purposes (*see* Section 2.3.2). Exposure to threatening, contradictory, or complex information is likely to increase levels of anxiety. A domain expert, however, is likely to use a schema-driven approach to information seeking (*see* Chapter 7). Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves without being driven by a need to reduce affective anxieties. This may help to explain how experts learn to tolerate uncertainty (Kuhlthau, 1993a; Kuhlthau, 1993b).

Wilson et al (2000) explored the difference in the overall level of uncertainty expressed by experiment participants with different levels of knowledge of their domain. Wilson et al found that as a participant's knowledge of the field increased so did the certainty expressed. Interestingly, the relationship between the level of uncertainty and the probable availability of relevant information was not significant. This finding is counter-intuitive. It was found that an individual's prediction of 'information availability' does not significantly affect levels of uncertainty. If an individual believes that they can find relevant information, it is very surprising that overall uncertainty is not reduced. This supports the idea that different types of uncertainty may exist (*see* Section 9.5.1).

To enable a better understanding, it is important to identify why an individual's prediction of 'information availability' does not, in general, reduce uncertainty. It is likely that a user's invitational or indicative mood results in a modification of information seeking heuristics (*see* Section 9.5.2). When an individual predicts 'information availability', uncertainty may not be reduced if that individual believes that the 'type of information' may not satisfy their current heuristic approach. Participants were asked to quantify their responses to the questions below on a magnitude scale. It was hypothesised that this would enable the predictive issue to be examined.

1) **"How much do you know about the retrieval task's subject domain?"**
   Not Enough ------------------------------>-------------------------- Enough

2) **"What type of information do you think you will need to complete the retrieval task?"**
   Background ---------------------------->-------------------------- Specific

The above questions were posed to participants after interaction with an information object. Question 1 enabled a self-assessment of domain knowledge to be made during the information seeking process. Question 2 enabled an individual to express an indicative or invitational mood.

9.5.6 TASK COMPLEXITY

The classification of problem stages encountered during information seeking enables the progress of individuals to be monitored. Exploring more subtle changes in a user's problem state (e.g., monitoring completion of sub-goals) requires a different approach. Search strategies for obtaining information and determining relevance are systematically connected to task complexity and the structure of the problem at hand. Vakkari (1999) believed that studies on information seeking behaviour have not contributed to how changes in the problematic situation and information needs are reflected in patterning of search strategies and relevance assessments.

Learning is most efficient when the structure to be learned is compatible with the structure that was expected according to prior knowledge (Heit, 1997). However, as Vakkari (1999) identified, one of the main difficulties in research has been how to describe changes in understanding, i.e., in the cognitive structures in a way that would connect the change process to changes in information actions. Heuristics are generated to enable an individual to address a problem in a way that is compatible with their learning abilities. Cognitive load does not allow an insight into the process of how an individual selects, apples, and modifies heuristics. However, understanding cognitive load may help to explain why individuals adopt a specific behavioural style.

The structure of the problem faced influences the patterning of strategies. Task complexity can be understood in many ways. Although, in general, an individuals ability to pre-determine task requirements governs the amount of complexity that the individual experiences. Pre-determining task requirements are related to the following characteristics of a task: repetitiveness, analysability, the number of alternative paths of task performance, and outcomes of novelty (Vakkari, 1999). The Geneplore model (*see* Section 9.2) was used to develop the passage feedback system used for experimentation (*see* Chapters 10 and 11). The feedback system allows specific types of cognitive processes to be explored, but also allows individuals to discover concepts that turn out to be much more important than the ones they were contemplating at the beginning. Repetitiveness, analysability, the number of alternative paths of task performance, and outcomes of novelty, can be investigated.

10.0 INTRODUCTION

The motivation for undertaking this experiment was to discover whether it is possible to show how an individual's cognitive limitations affect information behaviour. A mechanism is required that enables the degree to which a relevance feedback judgement has been inhibited by cognitive load to be evaluated.

Evaluating the usability associated with feedback capturing mechanisms requires the development of a Cognitive Load Theory. Educational psychologists originally developed the Cognitive Load Theory concept as a means of evaluating learning materials. The proposed extension to this theory within this thesis uses Cognitive Load Theory as a means of examining behaviour, rather than being concerned with suggesting techniques for minimising extraneous cognitive load (*see* Chapter 7).

A Cognitive Load Theory of information behaviour should have both an explanatory and predictive power. It is hoped that the flexibility of Cognitive Load Theory can be used to provide an improved understanding of behaviour that integrates both cognitive and affective dimensions.

Assumptions:

- Knowledge emerges as a result of a 'transfiguration process'. Information should not be viewed as a small part of a knowledge structure. Individuals purposefully construct meaning by selectively attending to that which connects with what they already know (*see* Chapter 2).
- An information need is an emergent realisation, triggered by the uncertainty associated with a gap in knowledge. The gap in knowledge may be between an individual's current state of domain knowledge and the perceived goal state, or between an individual's current state of procedural knowledge and the perceived goal state. Interactions that reduce uncertainty enable an insight into a user's state of knowledge (*see* Chapter 4).
- Sophisticated knowledge states can be held more easily in working memory. Less sophisticated knowledge states cannot be held entirely within working memory, so a higher cognitive load is imposed (*see* Chapter 7).

Conjectures (*see* Chapter 7 for theoretical postulate):

1. Cognitive load influences an individual's ability to interact with information objects.

2. Cognitive load experienced during relevance feedback is the intrinsic and extraneous difficulty associated with providing useful / accurate feedback judgements during information seeking.

3. More useful / accurate feedback judgements will be made if both an individual's intrinsic cognitive load and extraneous cognitive load is low.

4. Intrinsic load is the individual-specific difficulty associated with a retrieval task (dependent on user expertise and task complexity). Extraneous load is the individual-specific difficulty associated with identifying pertinent information (dependent on learner activity).

5. Both causal (intrinsic) and assessment (extraneous) factors affect cognitive load. Causal factors are the characteristics of an individual (e.g., cognitive abilities, and domain knowledge), while assessment factors are characteristics of the mental effort needed and the mental load imposed. Mental load is the portion of cognitive load that is imposed exclusively by the task and environmental demands.

6. Intrinsic factors are components of an individual's knowledge state, which is comprised of schemata, which are accessed using heuristics. Extraneous load is governed by the suitability of an individual's heuristics at processing information objects. More suitable heuristics enable information to be processed in a way which is more compatible with existing knowledge schemata.

10.1 AIMS, OBJECTIVES, AND HYPOTHESES

Aims:

1. To demonstrate whether the burden of cognitive load can be assessed within the context of providing relevancy judgements.

2. To demonstrate whether it is possible to encompass motivational and emotional processes within information processing models.

3. To demonstrate whether behavioural patterns are affected by cognitive load.

Objectives needed to fulfil Aim 1:

A. Identify which information seeking variables[33] can be used to predict the affect of intrinsic and extraneous cognitive load.

B. Model how these factors affect the utility of feedback judgements.

Objectives needed to fulfil Aim 2:

C. Determine the extent fluctuations in uncertainty during information seeking affects cognitive load.

D. Predict the extent manifestations of uncertainty influence information behaviour.

E. Explore the possibility that manifestations of uncertainty are symptomatic of limitations to cognitive capabilities.

Objective needed to fulfil Aim 3:

F. Evaluate the role heuristics play during information seeking.

---

[33] Factors to be assessed include: Knowledge obtained from previous search interactions, coping with information overload, exploratory v goal-directed behaviour, uncertainty, domain knowledge, and task complexity (*see* Section 9.5).

Hypothesis 1: Demographic Characteristics

There are no statistically significant relationships between:

I. Knowledge obtained from previous search interactions

II. Coping with information overload

III. Exploratory v goal-directed behaviour

IV. Uncertainty

V. Domain knowledge

VI. Task complexity

AND

A. the different sexes.

B. participants in different age groups.

C. participants from different occupational / academic backgrounds.

D. participants with different levels of on-line information retrieval experience.

E. participants with different levels relevance feedback experience.

Hypothesis 2: Characteristics of Predictor Variables

There are no statistically significant relationships between:

I. Knowledge obtained from previous search interactions

II. Coping with information overload

III. Exploratory v goal-directed behaviour

IV. Uncertainty

V. Domain knowledge

VI. Task complexity

AND

I. Knowledge obtained from previous search interactions

II. Coping with information overload

III. Exploratory v goal-directed behaviour

IV. Uncertainty

V. Domain knowledge

VI. Task complexity

Hypothesis 3: A Model of Intrinsic Cognitive Load (*see* Section 9.1)

There are no statistically significant relationships between:

I. Knowledge obtained from previous search interactions
II. Coping with information overload
III. Exploratory v goal-directed behaviour
IV. Uncertainty
V. Domain knowledge
VI. Task complexity

AND

$$\text{Predicted } imc = b_1X + b_2Y + \ldots + b_0$$

$$\text{Where } imc_j = \frac{icl}{bicl} = \frac{|iuj_j - puj_j|}{\left(\frac{puj_j}{tfe}\right)}$$

b_n = regression coefficients. b_0 = regression constant. *icl* = Intrinsic Cognitive Load. *bicl* = Baseline Intrinsic Cognitive Load. iuj_j = Initial Usefulness Judgement assigned to *j*. puj_j = Post-task Usefulness Judgement assigned to *j*. *tfe* = Task Fulfilment Evaluation.

Hypothesis 4: A Model of Extraneous Cognitive Load (*see* Section 9.3)

There are no statistically significant relationships between:

I. Knowledge obtained from previous search interactions
II. Coping with information overload
III. Exploratory v goal-directed behaviour
IV. Uncertainty
V. Domain knowledge
VI. Task complexity

AND

$$\text{Predicted } emc = b_1X + b_2Y + \ldots + b_0$$

$$\text{Where } emc = ecl = |ie_{pj} - pe_{pj}|$$

ecl = Extraneous Cognitive Load. ie_{pj} = Initial evaluation of passage usefulness for *pj*
pe_{pj} = Post-search evaluation of passage usefulness for *pj*

10.2 PARTICIPANTS

The mechanism underpinning this experimental procedure attempted to measure changes in cognitive load exclusively associated with the process of relevance feedback. An unfamiliarity associated with the process of information seeking may induce cognitive load. A within-subject experimental design was adopted in an attempt to enable cognitive load to be measured on individualistic comparative basis. Selecting a sample that was likely to possess generally comparable capabilities associated with IR increased the validity of the experimental design. Therefore, it was decided that participants should be experienced users of IR systems.

The decision to focus on individuals who use IR systems as a normal part of their everyday work activity was made. Individuals selected for experimentation were involved in an information intensive work environment. Participants were either involved in researching discursive academic areas or were information professionals. The main concern of the thesis was to discover if cognitive load is measurable on an individual basis. Individuals in information intensive work environments may have developed strategies for coping with cognitive load that are more explicitly observable. The intention was not to explore comparative differences associated with different groups of individuals. However, this may represent a future research direction.

Fifty individuals participated. Experimentation took place in May and August 2002. 20 participants performed the experiment in May. 30 participants performed the experiment in August. The same procedure was followed for both rounds of experimentation. However, automated response analysis software was written to expedite results obtained from the 2nd round of experimentation. Participants were recruited from five backgrounds: Five postgraduate psychologists at Nottingham University, five postgraduate human scientists at Loughborough University, five postgraduate information scientists at Loughborough University, five information professionals working for NHS Direct, five information professionals working for Midland Software, and 25 IR system developers / researchers at Autonomy. The age of participants ranged from 21 to 56. Their mean age was 29.480. There were 32 males and 18 females.

Selected participants had expressed an interest in helping with the development of advanced IR systems by responding to an e-mail. 72 individuals responded. 16 of these decided not to undertake the experimentation after being informed that it might take over one hour to complete. Six individuals were informed that they were not required due to difficulties associated with scheduling experimentation.

10.3 PROCEDURE

Exploratory Study 2 (*see* Section 8.1.1) used a method that allowed participants to select topic areas that they were likely to be interested in. Many of those participants (70%) suggested that the simulated retrieval tasks were interesting. The same method was therefore used here. Encyclopaedia Britannica provided a list of fifty of the most popular topic areas searched for in January 2002. This list was used to help generate a pool of seven tasks (*see* Appendix D). An email was sent to participants that included the list of the fifty topic areas, although participants were not informed of the source. Participants voted for five topic areas they were most interested in. These votes were collated and the top seven topic areas were used to generate the simulated tasks. This procedure attempted to minimise potential problems associated with participants undertaking tasks that they had no interest in.

Microsoft Visual Studio 6.0 was used to write the software required for this experimental procedure. A 'Microsoft Windows Web Browser' interface was developed. This type of interface was chosen on the basis that all participants selected for this experiment were familiar with opening / closing 'windows' and navigating Web pages. Standardised instructions were presented to participants alongside a questionnaire that gathered background data (*see* Figure 10.1). Participants were told that they were required to perform two tasks. Questions about how to use the software were permitted. Questions about the retrieval tasks were not permitted. Furthermore it was implied that if the participant became tired or disinterested, abandoning a task is an appropriate behaviour.

The following background variables were elicited for testing Hypothesis One: Sex [Variable A]; Age [Variable B]; Occupational or Academic background [Variable C]; On-line IR experience (magnitude scale 0-20) [Variable D]; Relevance feedback experience (magnitude scale 0-20) [Variable E] (*see* Section 10.1).

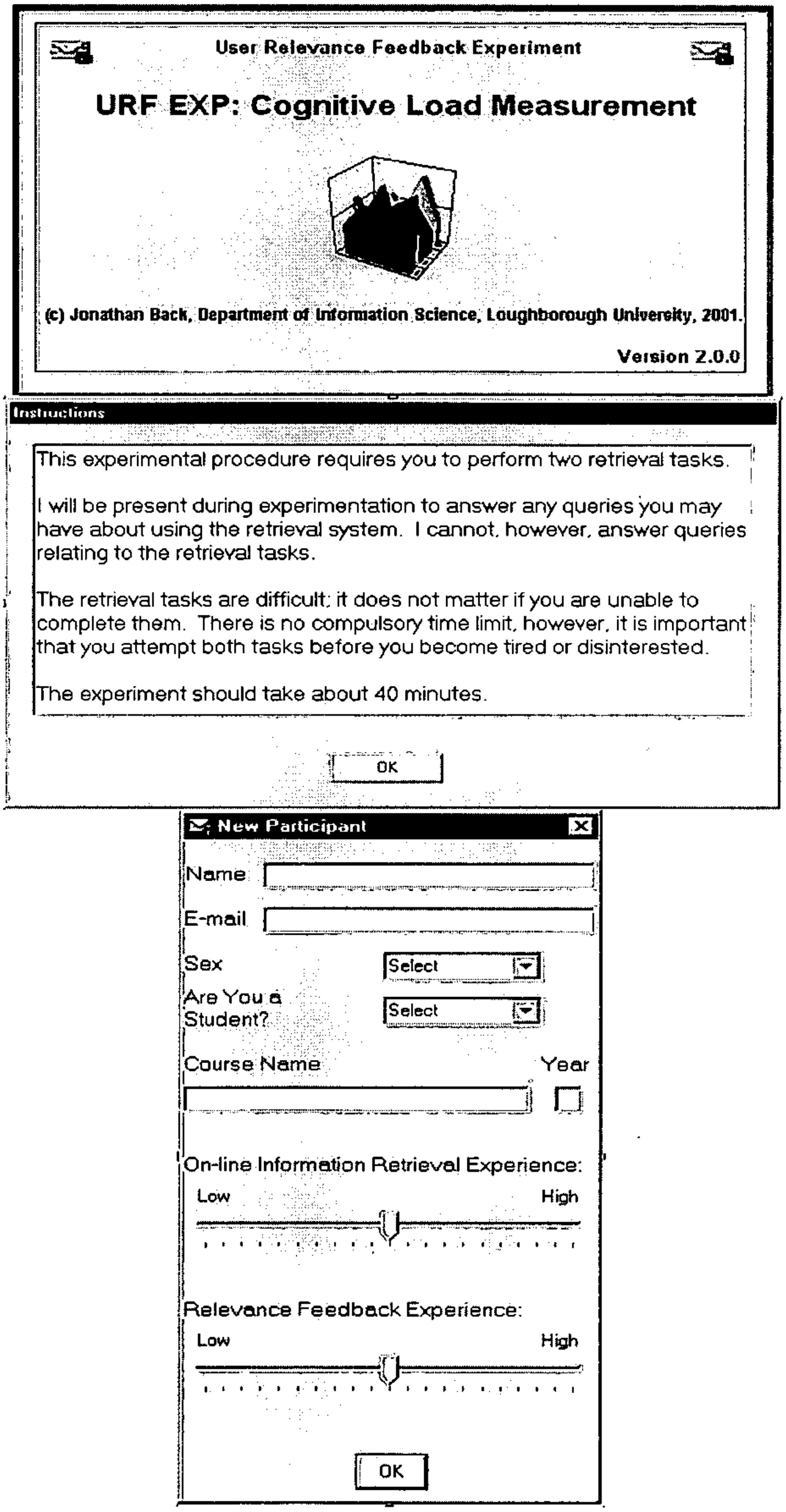


FIGURE 10.1: INSTRUCTIONS AND QUESTIONNAIRE

Participants were able to select which two tasks they wanted to perform from the pool of seven. Tasks were developed so that they were intricate/vague and often required a multipredicate hypothesis to be tested. The testing of multipredicate hypothesis often requires the critical evaluation of several different information objects. The intricate/vague nature of tasks encouraged users to form assumptions at the start of the task that may be refuted on completion. The simulated work task situation acts as the point of reference against which situational relevance is measured (i.e., the relationship between the information object and the user's underlying need). It can be suggested that the simulated work task provided freedom for each individual to react in relation to their own interpretation of the information need. An individual's interpretation of an information need is likely to require a combination of exploratory and goal-directed behaviour. Simulated work tasks encourage both these types of behaviour.

After selecting a task, participants were presented with an 'electronic notepad' (*see* Figure 10.2). They were informed that the 'notepad' was accessible at any time during the experimental process. The 'notepad' incorporated 'cut, copy, and paste' features. Clicking on the 'notepad' icon allowed the 'notepad' window to be opened and closed as required. Participants were told that the 'notepad' would ultimately be used to enable a written answer to the simulated task to be composed. However, at this stage, the notepad only contained details of task requirements.

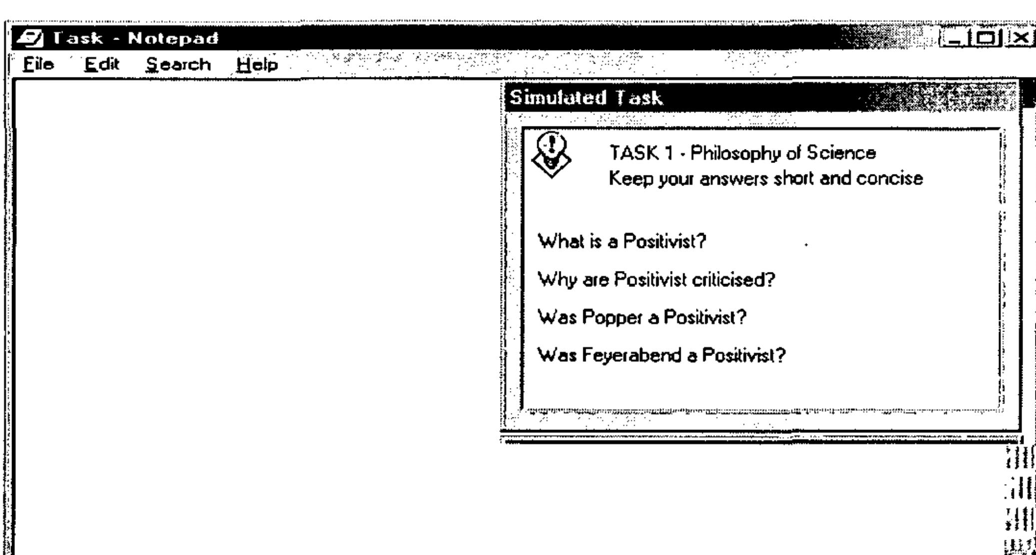


FIGURE 10.2: TASK NOTEPAD

10.3.1 INFORMATION BEHAVIOUR QUESTIONS

The next stage required participants to provide responses to four information seeking behaviour questions using magnitude scales (*see* Figure 10.3). None of the participants were familiar with answering information seeking behaviour questions of this type. An on-screen 'alert box' asked participants if they understood the questions. If they responded negatively, then a detailed verbal explanation was provided. The four questions allowed an assessment of a participant's expectations to be performed. Questions were posed before participants were permitted to interact with any information resources.

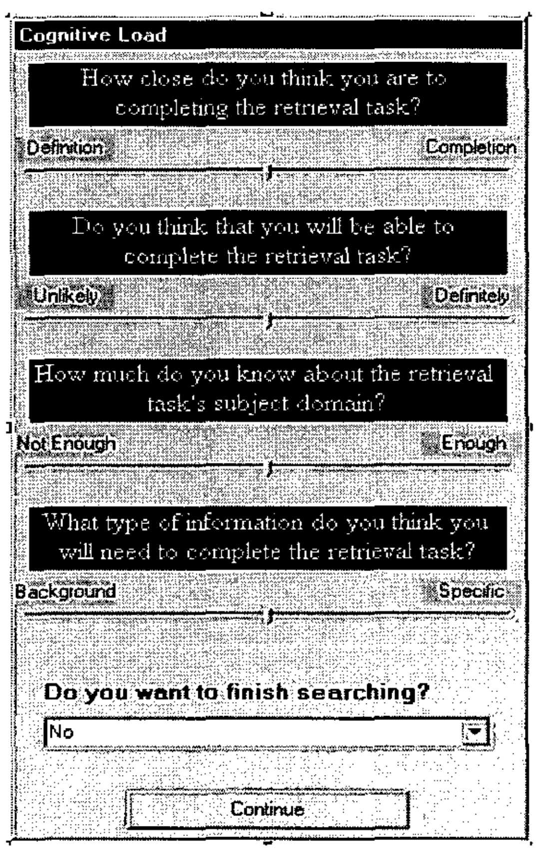


FIGURE 10.3: INFORMATION SEEKING BEHAVIOUR QUESTIONS

Question 1 "How close do you think you are to completing the retrieval task?", enabled the individual to self-report the sophistication of the knowledge schemata that he or she possessed. This question aimed to elicit the 'conceptual/cognitive state' uncertainty experienced by the individual at that specific point in time during information seeking (*see* Section 9.5.1) (Variable Label: **[con/cog_unc]**).

Question 2 "Do you think you will be able to complete the retrieval task?", allowed an individual to self-report the appropriateness of the problem-solving heuristics that they were currently applying or were considering applying. This question aimed to elicit the 'technical/affective' uncertainty experienced by an individual at that specific point in time during information seeking (*see* Section 9.5.1) (Variable Label: **[tec/aff_unc]**).

Question 3 "How much do you know about the retrieval task's subject domain?", enabled a self-assessment of domain knowledge to be made. A domain expert is likely to use a schema-driven approach to information seeking (*see* Chapter 7). Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves without being driven by a need to reduce affective anxieties (*see* Section 9.5.5). (Variable Label: **[dom_kno]**).

Question 4 "What type of information that do you think you will need to complete the retrieval task?", enabled an individual to express their indicative or invitational mood during information seeking. When an individual predicts 'information availability', uncertainty may not be reduced if that individual believes that the 'type of information' may not satisfy their current heuristic approach (*see* Section 9.5.5). (Variable Label: **[typ_kno]**).

10.3.2 RELEVANCE FEEDBACK PROCEDURE

Following the four questions being answered, the first document was displayed within a 'Web Browser' window. Information presented to experiment participants was cached from the Web version of the Encyclopaedia Britannica (permission was obtained). For each simulated task, pages were cached from the appropriate subject section of the Encyclopaedia Britannica. Each subject section of the online encyclopaedia has an introduction page. The first six linked documents on the introduction page were the ones cached. The introduction page was not cached. This forced participants to ascertain the conceptual links between the six documents for themselves (inducing cognitive load). Therefore, it was not always immediately obvious why documents were relevant for the task. If a linked document had its own links, then these links were deactivated. The presentation of documents, using this method, proved to be successful for Exploratory Study 2 (*see* Section 8.1).

The same method used in Exploratory Study 2 was adopted here since no confounding variables were found that could be associated with the idea that initial and post-task judgements may improve or worsen as more documents are presented. The documents presented to volunteers varied in length from 500 - 2000 words. Participants were encouraged to browse documents rather than read them thoroughly. Participants spent, on average, four minutes per document. Information was technical but concepts were presented in an entertaining way rather than being terse. The six documents presented for each task were all topically very relevant. However, as the introductory page was not cached and the six documents were presented randomly, the likelihood of fluctuations in individualistic situational relevance occurring was high.

Participants were informed that they could revisit a page if they wanted to, but had to attempt a relevance feedback procedure 'first time round' even if they found it hard to do so. A user may well be unable to communicate their cognitive state to a system. A user's perceived mental model may not be an accurate reflection of their information need situation. The relevance feedback procedure involved selecting any passages that were considered to be potentially 'useful' using a highlighting tool. Ideally, usefulness judgements should be based on an individual's perception of their information problem. It was anticipated that each participant would interpret the work task differently to simulate a 'cognitive breakdown situation'. It was important that judgements related to the participant's mental state at the time of being presented with a Web page. Participants were informed that usefulness judgements should be 'their own opinion of how important the information is'.

The feedback procedure adopted here was similar to the RF_b mechanism used in Exploratory Study 1 (*see* Chapter 3). Both the cognitive and situational factors affecting usefulness are continually fluctuating. Ideally, a feedback mechanism should be designed in a way that enables spontaneous judgments to be captured. Therefore, feedback was submitted by depressing the left mouse button and moving the mouse cursor to highlight useful passages. This enabled the individual semantic entities of a document (sections, paragraphs, sentences, etc) to be used to generate a usefulness judgment. Results from Exploratory Study 1 showed that using a highlighting tool increased the likelihood of more useful judgments being made.

On submitting a passage for feedback purposes, two questions were asked (*see* Figure 10.4). These questions evaluated a participant's level of understanding. The process of retrieving existing structures from memory is analogous to selecting a passage from a document for feedback purposes. It can be assumed that the reason why a participant selected a passage was because it had a novel property, which in turn facilitated the finding of attributes, which enabled the creation of conceptual combinations that were needed for associations to be made (*see* Section 9.2). The process of association was investigated by asking a participant if they "fully understand the contextual meaning of the selected passage?" (Question 2) (Variable Label: **[context]**). It can be assumed that if a participant suggested that they did not fully understand, then this reveals ambiguity. Conceptual interpretation is likely to be triggered by ambiguity, which in turn may enable synthesis.

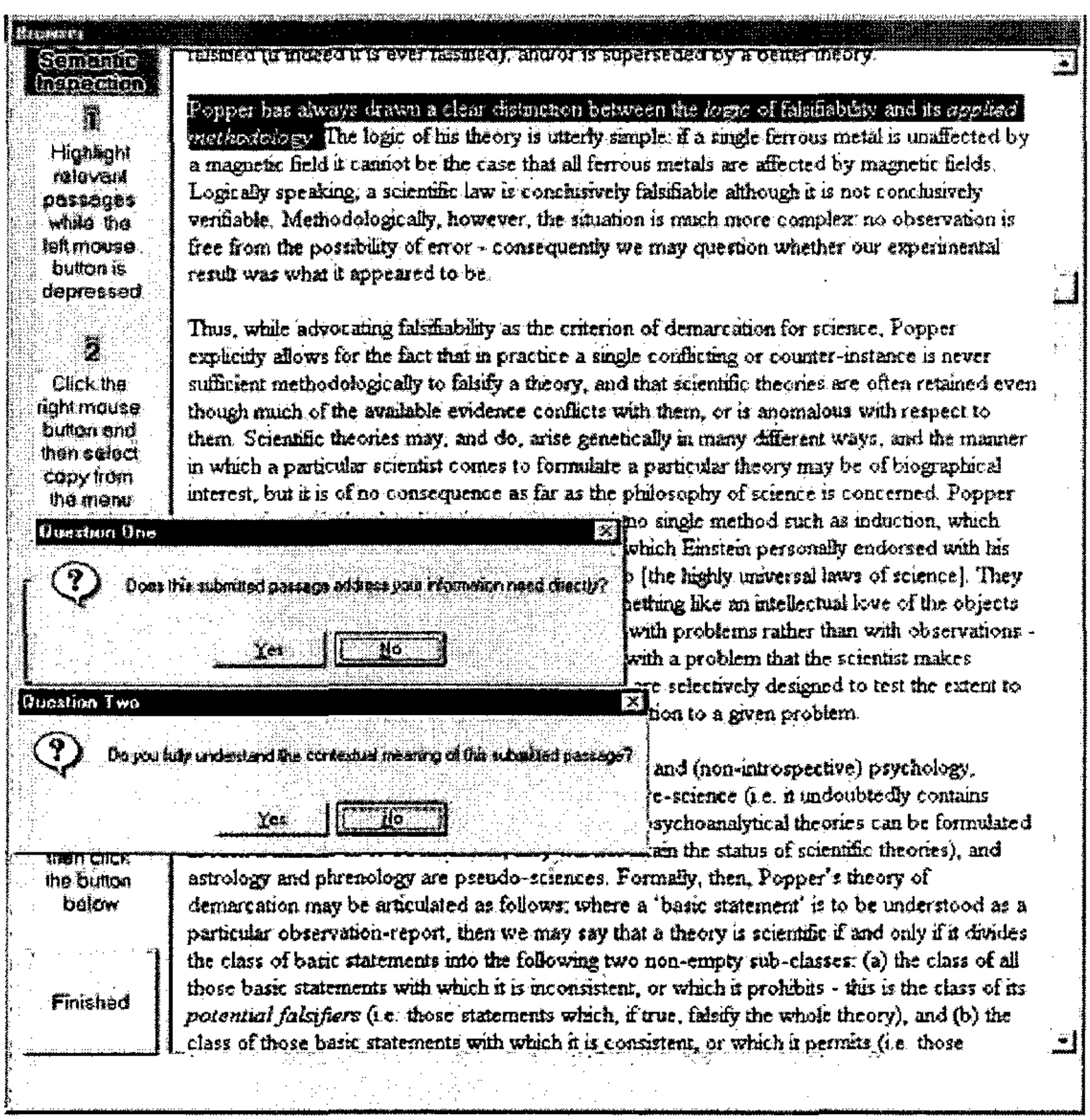


FIGURE 10.4: RELEVANCE FEEDBACK PROCEDURE

Question 1 "Does the submitted passage address your information need directly?", allowed the process of synthesis to be investigated (Variable Label: **[blend]**). If a participant was able to perform an evaluation of utility, this indicated that a *mental blend* had occurred (as the evaluation of a specific passage is performed with respect to all other passages and existing knowledge). If a passage was deemed only to be potentially useful, then this indicated that a search for meaningfulness was underway.

Once the passage feedback procedure had been completed, participants were required to evaluate the overall usefulness of the Web page using a magnitude scale (*see* Figure 10.5). This procedure enabled the 'Initial Usefulness Judgement' to be assigned (Variable Label: **[iuj]**). Eliciting this variable was required for the calculation of intrinsic cognitive load (*see* Section 9.1). Two 'check-box' options were also presented: 'I need to browse documents that are of a similar nature to this one' (Variable Label: **[similar]**); 'I need to browse documents that are substantially different to this one' (Variable Label: **[different]**). Eliciting these variables helped elucidate possible exploratory or goal-directed behavioural intentions.

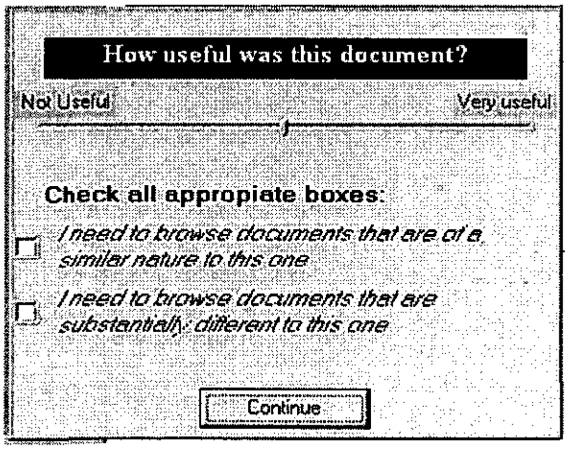


FIGURE 10.5: INITIAL USEFULNESS JUDGEMENT

Crucially, participants were then returned to the four information seeking behaviour questions (*see* Figure 10.3). A re-evaluation of responses using the magnitude scales enabled an assessment of how a task was progressing. These responses were matched with the 'Initial Usefulness Judgement' score assigned to the previous document browsed. This allowed the situational context, in which the 'Initial Usefulness Judgement' was made, to be revealed.

10.3.3 ANSWER CONSTRUCTION

When investigating extraneous cognitive load, a mechanism was needed that allowed changes in an individual's heuristics during the process of information seeking to be discovered (*see* Section 9.3). After re-evaluating the information seeking behaviour questions, participants were presented with the 'electronic notepad' (*see* Figure 10.2). Participants were required to construct a written answer to the simulated retrieval task. The answer construction process enabled conceptual elements, from passages submitted for feedback purposes, to be incorporated by participants. Participants had to provide a written answer or they failed to complete the task. Participants were not informed of the number of documents presented for each task; this encouraged them to attempt an answer after browsing each document. Each attempt at an answer was representative of the heuristics applied at a specific point in time. Heuristics change over time and the notes made by an individual may be invalidated or become less useful when an individual applies modified heuristics. This allowed extraneous cognitive load to be evaluated.

Passages submitted for feedback purposes were presented back to participants. Participants were encouraged to attempt to incorporate these passages within their answer. Answer construction is a dynamic process. As the information seeking process progresses, participants were able to continually update and re-examine their answer. The number of times feedback passages could be removed, modified, and re-incorporated during answer construction was unlimited. Participants could click on the 'notepad' icon at any point during experimentation.

Although adding complementary words/sentences was expected, it was advised that it would be preferable to attempt to construct an answer that made as much use of feedback passages as possible. The purpose of the first phase of the answer construction process was to allow the calculation of an 'Initial evaluation of passage usefulness' measure (Variable Label: [**ie**]). This first phase involved assigning either a *red* or *yellow* flag to submitted passages, or alternatively, being able to directly incorporate passages into an answer construction (*see* Figure 10.6).

Red Flag: Participants were asked to identify if a feedback passage was already conceptually similar to the existing answer construction, and therefore not suitable for incorporation. This was achieved by assigning a *red flag* to the passage. Submitted passages assigned a *red flag* were given an *ie* score of 1. An *ie* score of 1 indicated that a submitted passage was, initially, not useful for answer construction.

Yellow Flag: Participants were asked to identify if a feedback passage was potentially useful. This was achieved by assigning a *yellow flag* to the passage. Submitted passages assigned a yellow flag were given an *ie* score of 2. An *ie* score of 2 indicated that a submitted passage was, initially, considered to be potentially useful for the answer construction process at a later time.

Green Flag: If participants used a submitted passage for the answer construction process immediately, an *ie* score of 3 was assigned. An *ie* score of 3 indicated that a passage was, initially, highly useful.

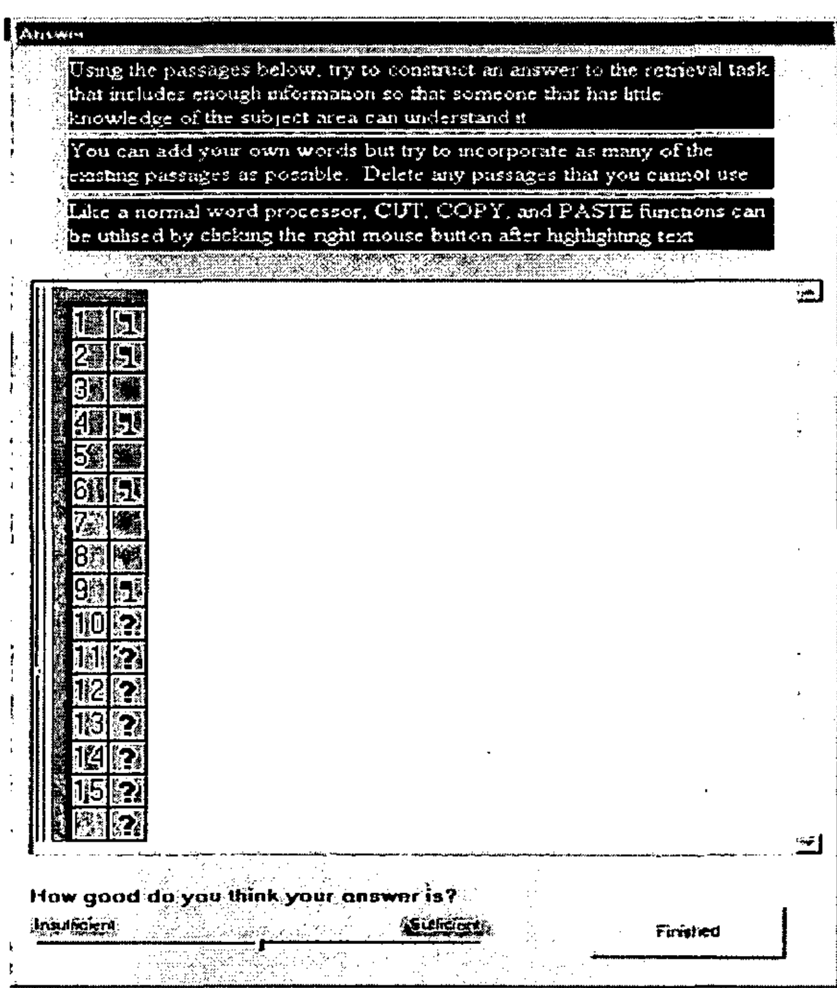


FIGURE 10.6: ANSWER CONSTRUCTION

10.3.4 POST-TASK EVALUATION

Figure 10.7 outlines the experimental procedure adopted. Assigning feedback judgements (*see* Section 10.3.2), answering information seeking behaviour questions (*see* Section 10.3.1), and performing the answer construction process (*see* Section 10.3.3), had to be performed after every document browsed. When the final answer construct was composed, a post-task evaluation procedure was performed. Post-task evaluation is essential for the development of intrinsic and extraneous measures of comparison. These measures were used to assess the cognitive load imposed on an individual during information seeking (*see* Sections 9.1 and 9.3).

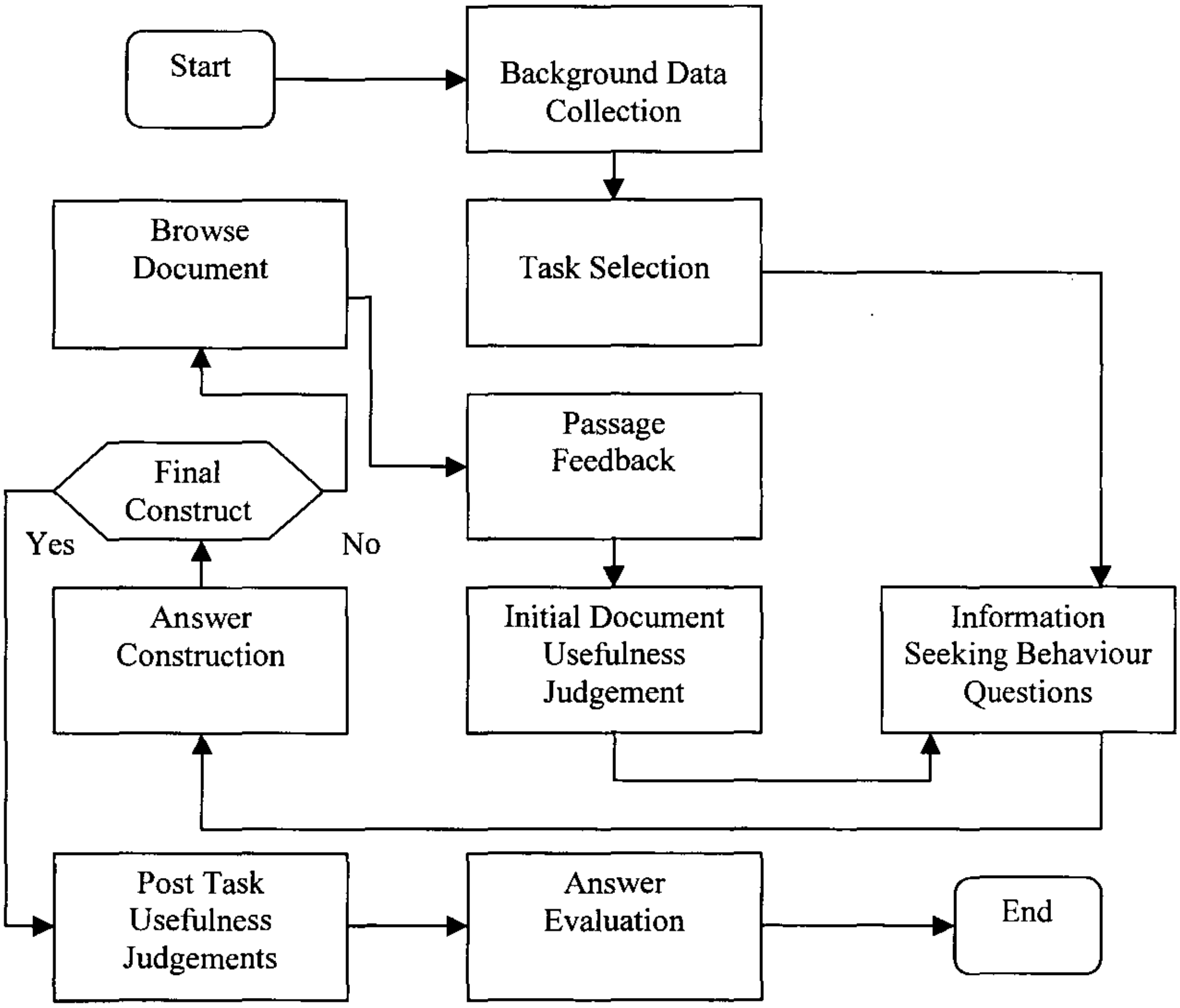


FIGURE 10.7: EXPERIMENTAL PROCEDURE

All documents browsed were sequentially presented back for a post-task usefulness evaluation using a magnitude scale. This enabled the capture of the 'Post-task Usefulness Judgement' measure (Variable Label: [**pu**]) (*see* Section 9.1). The scale was identical to the one used to elicit the 'Initial Usefulness Judgement' (*see* Figure 10.5). Participants were instructed that the usefulness judgement should be 'their own opinion of how important the information was to performing the task'.

During experimentation, the intrinsic cognitive load imposed on an individual was calculated for each individual document (information object). A comparative analysis of intrinsic cognitive load measures across a range of tasks/participants could not be performed without normalisation. Therefore, a baseline task performance normalisation was performed. Participants were required to assess whether they had sufficiently addressed the retrieval task by using a magnitude scale (Variable Label: [**tfe**]). This scale was presented alongside a participant's answer construction (*see* Figure 10.6). The question "How good do you think your answer is?" was posed. The required answer was elicited on an 'insufficient' to 'sufficient' magnitude scale. Participants were not permitted to make changes to their answer.

10.4 VARIABLES CAPTURED FOR HYPOTHESES TESTING

The procedure proposed above allowed a quantitative investigation of ideas developed through qualitative research. All the variables elicited aimed to provide a quantitative insight into behavioural processes. It must be emphasised that eliciting predictor variables (identified by qualitative research), that aim to satisfy a mathematical model, may only provide a limited insight. **Every interaction performed by a participant was logged. It was essential that behavioural patterns could be observed during the process of answer construction.** Many of the variables needed to develop a model of cognitive load can only be assessed by identifying 'characteristic patterns' of behaviour.

The following variables were self-reported by participants during the experimentation:

Sex **[sex]**, Age **[age]**, Occupational / academic backgrounds **[back]**, Level of on-line Information retrieval experience **[ir]**, Level relevance feedback experience **[rf]**. Conceptual / cognitive state uncertainty **[con/cog_unc]**, Technical / affective uncertainty **[tec/aff_unc]**, Domain knowledge **[dom_kno]**, Indicative or invitational mood **[typ_kno]**. Contextual meaning of the selected passage **[context]**, Extent passage addresses information need **[blend]**, Initial usefulness judgement **[iuj]**, Goal-directed intentions **[similar]**, Exploratory intentions **[different]**. Initial evaluation of passage usefulness **[ie]**. Post-task usefulness judgement **[pu]**, Task fulfilment evaluation **[tfe]**.

Variables needed to test all hypotheses (see Section 10.1):

Background Variables:

A. Sex. **[sex]**

B. Age. **[age]**

C. Occupational / academic background. **[back]**

D. Levels of on-line information retrieval experience. **[ir]**

E. Levels relevance feedback experience. **[rf]**

Predictor Variables:

I. Knowledge from previous interactions. **[con/cog_unc] & [tec/aff_unc]**

II. Information overload. **[typ_kno] & [Behavioural Characteristics]**

III. Exploratory / goal-directed **[similar] & [different] & [Behavioural Characteristics]**

IV. Uncertainty. **[con/cog_unc] & [tec/aff_unc] & [Behavioural Characteristics]**

V. Domain knowledge. **[dom_kno]**

VI. Task complexity. **[Behavioural Characteristics]**

Intrinsic and Extraneous Measures of Comparison:

1. Initial Usefulness Judgement. **[iuj]**
2. Post-task Usefulness Judgement. **[pu]**
3. Task Fulfilment Evaluation **[tfe]**
4. Initial evaluation of passage usefulness. **[ie]**
5. Post-search evaluation of passage usefulness. **[Not self-reported. *See* Section 9.3]**

11.0 INTRODUCTION

Experimentation enabled 72 hypotheses and six conjectures to be tested (*see* Section 10.1). 50 participants undertook two tasks from a pool of seven. In total; 600 intrinsic and extraneous measures of cognitive load were calculated (12 for each participant); 60 different Web pages were browsed (12 for each participant); 2,468 passage relevance feedback judgements were made (mean = 49.36 for each participant); 5,459 interactions occurred during the answer construction process (mean = 109.18 for each participant); experiment time was 63.67 hours (mean = 76.40 minutes per participant).

11.1 RELIABILITY

A number of tests can be performed to determine whether the results obtained from experimentation are generalizable. All participants claimed they understood the principle of assigning usefulness judgements to documents and providing passage feedback judgements. Figures 11.1 and 11.2 show that initial and post-task usefulness judgements assigned to documents did not significantly improve or worsen as more documents were presented for evaluation. A linear regression was used to verify this. These findings suggest that the usefulness of a document can be assessed on an individual basis, independently of other documents evaluated. It was also found that a participant's age, gender, occupational background, online search experience, and relevance feedback experience did not significantly affect results.

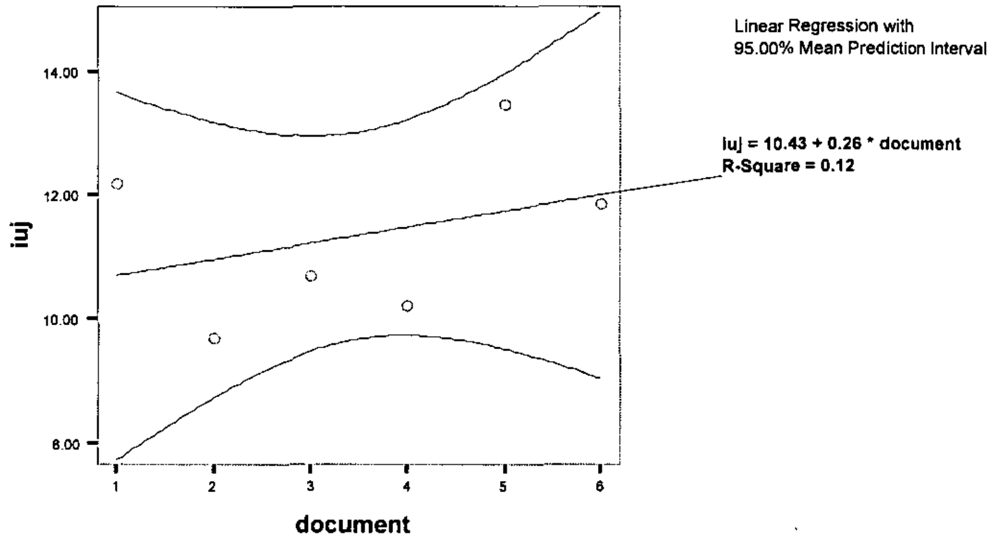


FIGURE 11.1: INITIAL USEFULNESS JUDGEMENT RELIABILITY

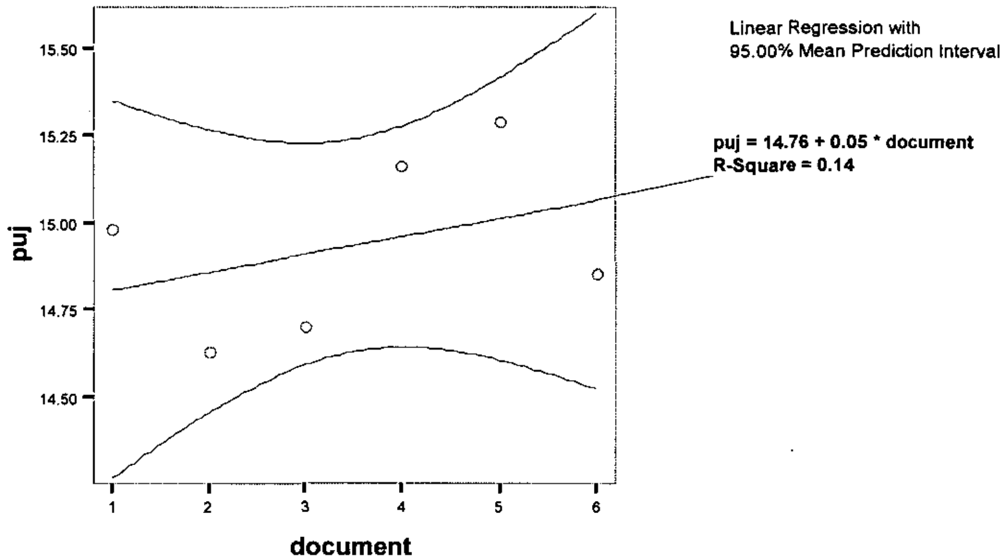


FIGURE 11.2: POST-TASK USEFULNESS JUDGEMENT RELIABILITY

Figure 11.3 shows that a post-task judgement assigned to a document is not significantly related to the initial judgement made. A usefulness judgement can, therefore, be considered as a measure of situational needs. If a significant correlation existed, this would indicate that usefulness might be a surrogate measure for the increased confidence experienced by a participant on completion of a task.

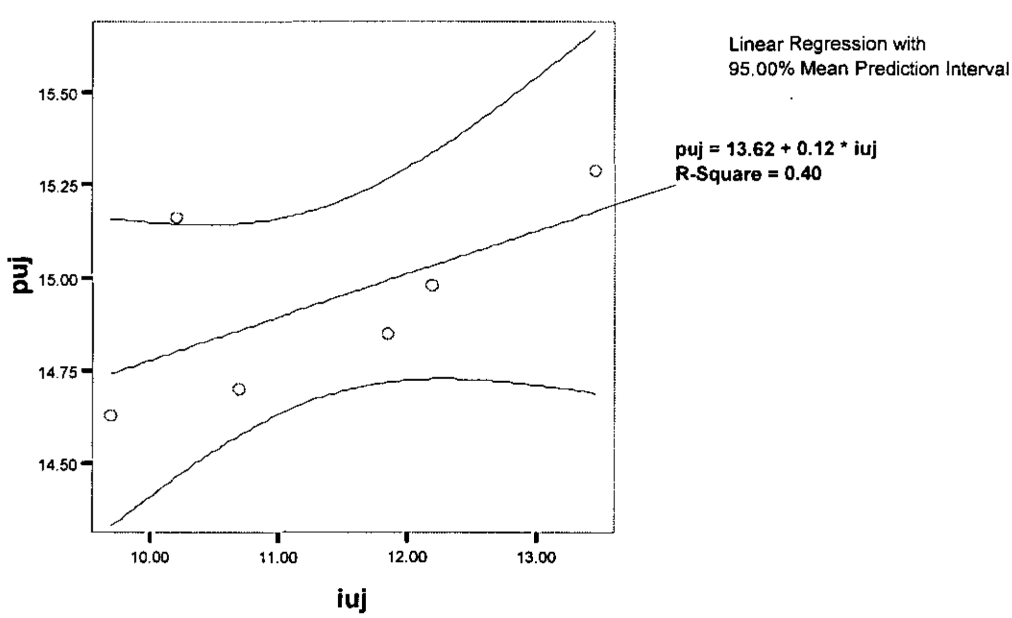


FIGURE 11.3: INITIAL V POST USEFULNESS JUDGEMENT RELIABILITY

The initial evaluation of passage usefulness (*ie*) involved assigning either a *red* (*ie*=1) or *yellow* (*ie*=2) flag to submitted passages, or alternatively, being able to directly incorporate passages into an answer construction (*ie*=3). Figure 11.4 shows the mean *ie* score assigned to passages submitted from each of the six documents. It can be seen that the mean *ie* score does not significantly improve or worsen as more documents were browsed. These findings suggest that the initial evaluation of passage usefulness can be performed on an independent basis, without being influenced by browsing more documents. It was also found that a participant's age, gender, occupational background, online search experience, and relevance feedback experience did not significantly affect results.

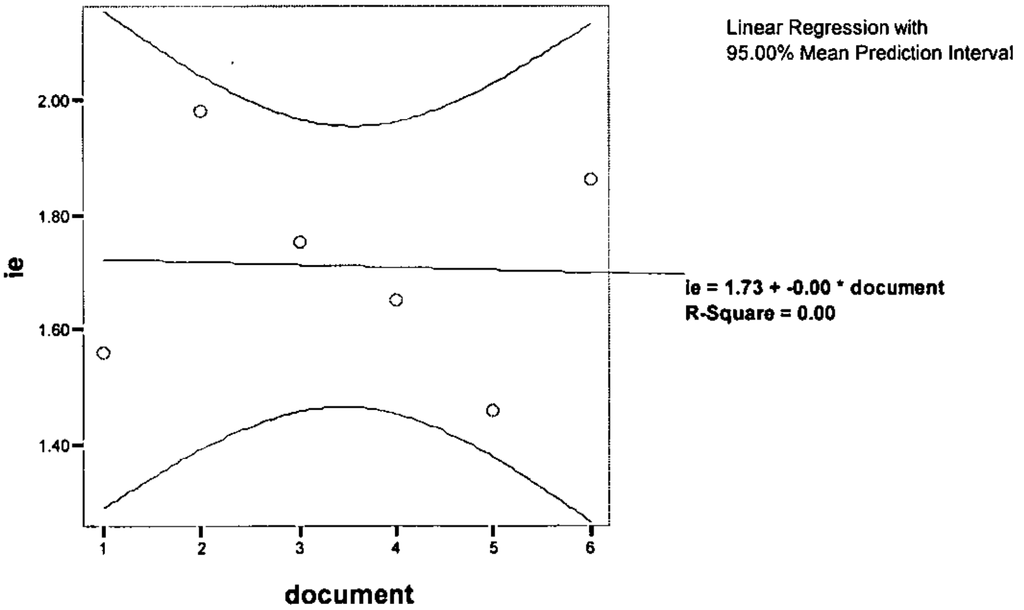


FIGURE 11.4: INITIAL EVALUATION OF PASSAGE USEFULNESS RELIABILITY

Figure 11.5 shows that a post-task passage usefulness score (*see* Section 9.3) is not significantly related to the initial evaluation of passage usefulness. A passage usefulness judgement can, therefore, be considered as a measure of situational needs. If a significant correlation existed, this would indicate that passage usefulness might be a surrogate measure for the increased levels of knowledge possessed by a participant on completion of a task.

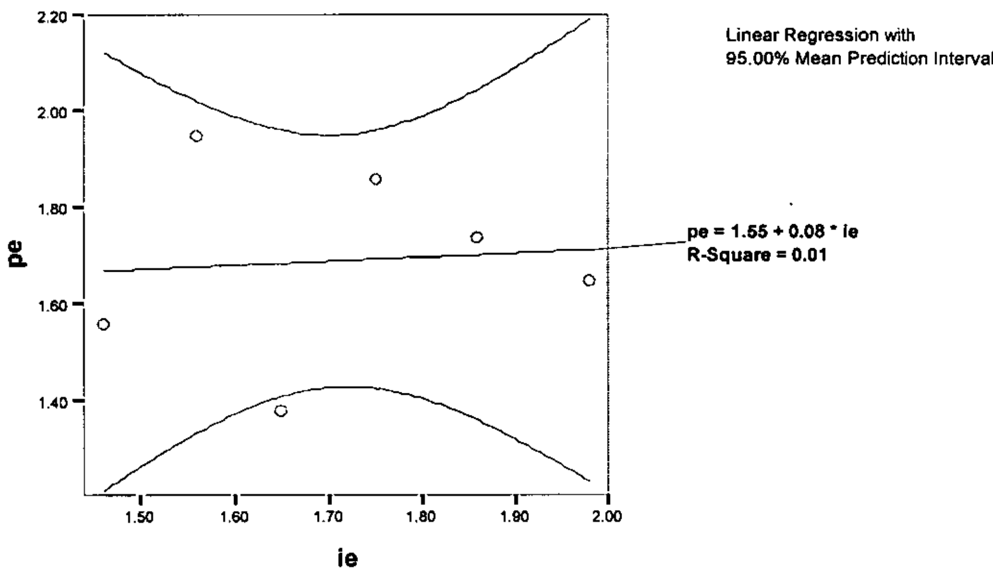


FIGURE 11.5: INITIAL V POST PASSAGE USEFULNESS RELIABILITY

Figures 11.1, 11.2, 11.3, 11.4 and 11.5 show that the method adopted to enable the calculation of intrinsic and extraneous measures of comparison is reliable (*see* Sections 9.1 and 9.3). Measures of comparison are not influenced by confounding variables. Explanations for the variance between initial and post-task usefulness evaluations are attributable to an individual's situational context.

11.2 TASK FULFILMENT

Tasks were generally well received. Many participants (78%) suggested that they were interesting. Table 11.1 shows how participants self-evaluated task fulfilment. Categorisation of a participant's task success was based on a score elicited on a magnitude scale from 0-20. A high level of residual intrinsic cognitive load is inevitable if task performance is poor. A low level of residual intrinsic load is expected when task performance is good. A baseline measure enables intrinsic measures of comparison to be normalised (*see* Section 9.1). 100 Tasks were performed. It was decided that a self-evaluated score of 15 or over required no baseline as it can be anticipated that residual cognitive load was minimal. A score of 10-14 was assigned a baseline of 1.22: some residual load was anticipated. A score of 6-10 was assigned a baseline of 1.38: significant residual load was anticipated. No task was assigned a score lower than six.

TASK FULFILMENT	TASKS	MEAN INITIAL V POST-TASK PASSAGE UTILITY VARIANCE	
15 or Over	43	0.985	
10-14	39	0.764 (-22.437%)	Baseline = 1.22
6-10	18	0.598 (-37.578%)	Baseline = 1.38
TOTAL	100		

TABLE 11.1: TASK SUCCESS RATE

'Mean initial v post-task passage utility variance' (*see* Section 9.3) was used to calculate the values of the baseline measures. Table 11.1 (above) shows that participants who assigned a higher task fulfilment score experienced a greater degree of passage utility variance. During the answer construction process participants who made more modifications were likely to be engaged in a more active learning / problem-solving process (*see* Section 11.10). This indicates that the cognitive load imposed on them was not high enough to prevent the modification of heuristics (*see* Section 9.2). In summary, it can be suggested that the tasks used for experimentation were challenging. Participants who performed better in tasks were able to constantly re-evaluate their information need situation.

11.3 KNOWLEDGE OBTAINED FROM PREVIOUS SEARCH INTERACTIONS

- Q1: "How close do you think you are to completing the retrieval task?"
- Q2 "Do you think you will be able to complete the retrieval task?"

It was anticipated that Q1 might allow 'conceptual' and 'cognitive state' uncertainty to be measured at a specific point in time during information seeking (*see* Section 9.5.1). It was suggested that individuals with well-developed knowledge schemata enable both 'conceptual' and 'cognitive state' uncertainty to be lowered. Q2 aimed to elicit the 'technical' / 'affective' uncertainty experienced by an individual at a specific point in time during information seeking. It was anticipated that Q2 allowed an individual to self-report the appropriateness of the heuristics that they were applying.

Behavioural manifestations observed during the answer construction process (*see* Section 10.3.3) provided evidence to support the assumptions that Q1 and Q2 were reliable measures. Behaviours observed included: the number of feedback passages assigned a red flag (*ie=1*), and the number of feedback passages assigned a yellow flag (*ie=2*). Table 11.2 shows that a significant negative Pearson's correlation between the number of passages assigned *ie* = *1* and Q1 existed. Individuals increasingly identified which passages were not useful when they moved closer to task completion. Table 11.3 shows that a significant positive Pearson's correlation between the number of passages assigned *ie* = *2* and Q2 existed. An ability to identify which passages may be potentially useful perpetuated an individual's uncertainty associated with completing a task.

Correlations

		Q1	ie1
Q1	Pearson Correlation	1.000	-.702*
	Sig. (2-tailed)	.	.011
	N	600	600
ie1	Pearson Correlation	-.702*	1.000
	Sig. (2-tailed)	.011	.
	N	600	600

*. Correlation is significant at the 0.05 level (2-tailed).

TABLE 11.2: NOT USEFUL PASSAGES

Correlations

		Q1	ie2
Q2	Pearson Correlation	1.000	.817**
	Sig. (2-tailed)	.	.001
	N	600	600
ie2	Pearson Correlation	.817**	1.000
	Sig. (2-tailed)	.001	.
	N	600	600

**. Correlation is significant at the 0.01 level

TABLE 11.3: POTENTIALLY USEFUL PASSAGES

A participant's age, gender, occupational background, online search experience, and relevance feedback experience did not significantly affect reported Q1 and Q2 levels. Therefore, **Hypothesis 1 (I: A, B, C, D, E)** is supported (*see* Section 10.1).

11.4 COPING WITH INFORMATION OVERLOAD

This thesis has avoided making the assumption that uncertainty is a suitable measure of an individual's ability to process information (*see* Section 9.5.2). The possibility that manifestations of uncertainty are symptomatic of limitations to cognitive capabilities needs to be considered (*see* Section 11.6). When evaluating how individuals cope with information overload, it is important to discover if a user's mood (invitational or indicative) results in a modification of information seeking behaviour (*see* Section 10.3.2).

Behaviour observed during the passage feedback process failed to provided evidence to support the assumption that a user's invitational or indicative mood was a reliable measure. It can be assumed that the reason why a participant selects a passage is because it has a novel property, which in turn facilitates the finding of attributes, which may enable the creation of conceptual combinations that are needed for associations to be made (*see* Section 9.2). The process of association can be investigated by asking a participant if they "fully understand the contextual meaning of the selected passage?". Table 11.4 shows that no significant correlation exist between a user's mood and responses to this question. **Hypothesis 1 (II: A, B, C, D, E)** cannot be tested, as a user's mood is not a suitable measure for how an individual copes with information overload (*see* Section 10.1).

Correlations

		context	mood
context	Pearson Correlation	1.000	-.300
	Sig. (2-tailed)	.	.121
	N	600	600
mood	Pearson Correlation	-.300	1.000
	Sig. (2-tailed)	.121	.
	N	600	600

TABLE 11.4: MOOD V UNDERSTANDING

11.5 EXPLORATORY V GOAL-DIRECTED BEHAVIOUR

An individual's investigation of an information need is likely to require a combination of exploratory and goal-directed behaviour. It would be interesting to discover how exploratory or goal-directed behaviour affects cognitive load. A measure needs to be found that allows exploratory / goal-directed behaviour to be represented within a model of cognitive load. The answer construction process enabled both exploratory and goal-directed behaviours to be reflected. Exploratory behaviour could be characterised by participants being less willing to combine concepts during the answer construction process. Goal-directed behaviour could be observed when participants decided to integrate concepts. Post-task passage evaluation scores (*pe*) were used to examine this behaviour (*see* Section 9.3). If participants used a submitted passage for the answer construction process and modified the semantics of the passage, a $pe = 2$ score was assigned. A $pe = 2$ score indicated that a passage had been integrated in some way with one or more different concepts. This is symptomatic of goal-directed behaviour.

Two 'check-box' options were presented at the usefulness evaluation stage of the experimental procedure (*see* Section 10.3.2): 'I need to browse documents that are of a similar nature to this one' (Variable Label: [**similar**]); 'I need to browse documents that are substantially different to this one' (Variable Label: [**different**]). It was suggested that these 'check-box' options might help to elucidate behavioural intentions. No significant relationships were found associated with the 'different' checkbox. However, Table 11.5 shows that when a $pe = 2$ assignment was made, a significantly higher proportion of 'check-box ticks' associated with the requirement to browse similar documents was found (sig. at the 0.01 level).

PE ASSIGNMENT	SIMILAR CHECK-BOX TICKS	PERCENT
$pe = 0$	57	24.359%
$pe = 1$	18	7.692%
$pe = 2$	129	55.128%
$pe = 3$	30	12.820%
Total	234	100.000%

TABLE 11.5: GOAL-DIRECTED BEHAVIOUR

A suitable measure of an individual's goal-directed / exploratory behaviour has been found. The *pe=2* measure can be reliably used within a model of cognitive load. It was found that a participant's age, gender, occupational background, online search experience, and relevance feedback experience did not significantly affect *pe=2* values. Therefore, **Hypothesis 1 (III: A, B, C, D, E)** is supported (*see* Section 10.1).

11.6 UNCERTAINTY

Exploring the extent fluctuations in uncertainty during information seeking affects cognitive load needs to be determined. If dissonance occurs during IR interaction, then it is unlikely that behaviour will change unless the strength of the dissonance increases a user's level of uncertainty to a level that affects the cognitive load imposed on working memory (*see* Section 9.5.4). Testing, reported in Section 11.3, found that individuals can increasingly identify which passages are not useful when they move closer to task completion. Furthermore, an ability to identify which passages may be potentially useful was found to perpetuate an individual's uncertainty associated with completing a task.

A stepwise regression analysis (*see* Table 11.6) revealed that an 'Initial Utility Judgement' could be predicted using two factors (standard error of the estimate = 2.20). The number of passages identified by a participant as being not useful was significantly positively correlated (Pearson correlation = 0.703). The number of passages identified by a participant as being potentially useful was significantly negatively correlated (Pearson correlation = -0.629). If a participant was able to identify passages as not being useful, this implied that 'conceptual and cognitive state' uncertainty was low enough to allow this behaviour. If a participant was able to identify passages as being potentially useful, but was unable to integrate these within an answer construction, this perpetuated 'technical / affective' uncertainty. Results show that an 'Initial Utility Judgement' can be better explained when both 'conceptual and cognitive state' and 'technical / affective' uncertainty were considered at the same time. A higher 'Initial Utility Judgement' is predictable when both types of uncertainty are low. A lower 'Initial Utility Judgement' is predictable when both types of uncertainty are high. Both measures of uncertainty are suitable for the inclusion within a model of cognitive load. Although they measure separate elements, when combined, they enable an assessment of overall uncertainty.

INITIAL USEFULNESS JUDGEMENT ASSIGNMENT

Model 1: Number of passages identified by a participant as being not useful.
Model 2: Number of passages identified as being potentially useful + Model 1

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	2.55	.495	9.797	.011
2	2.20	.165	4.379	.066

TABLE 11.6: DERIVING A MODEL OF UNCERTAINTY

A participant's age, gender, occupational background, online search experience, and relevance feedback experience did not significantly affect the number of passages identified by a participant as being not useful, or the number of passages identified by a participant as being potentially useful. Therefore, **Hypothesis 1 (IV: A, B, C, D, E)** is supported (*see* Section 10.1).

11.7 DOMAIN KNOWLEDGE

Exposure to threatening, contradictory, or complex information is likely to increase levels of anxiety. A domain expert, however, is likely to use a schema-driven approach to information seeking (*see* Chapter 7). Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves without being driven by a need to reduce affective anxieties. This may help to explain how experts learn to tolerate uncertainty. When an individual predicts 'information availability', uncertainty may not be reduced if that individual believes that the 'type of information' may not satisfy their current heuristic approach. Questions 3 and 4 aimed to elicit answers that enabled these issues to be examined.

- Q3: "How much do you know about the retrieval task's subject domain?"

- Q4: "What type of information do you think you will need to complete the retrieval task?"

It was anticipated that Q3 would enable a self-assessment of domain knowledge to be made during the information seeking process, and Q4 would enable an individual to express their indicative or invitational mood. The aim was to show if a domain expert is better at tolerating 'technical / affective' uncertainty (Q2 *see* Section 11.3).

A stepwise regression analysis (*see* Table 11.7) revealed the factors that were able to provide an insight into how 'technical / affective uncertainty' (Q2) was self-reported by participants. Results show that Q4, 'the extent an information need has been addressed' (Variable Label: **[blend]**) (*see* Section 10.3.2), and Q1, were excellent predictors (standard error of the estimate = 0.71).

- When a participant suggested that they required specific information to complete a task (Q4), they were frequently experiencing lower levels of 'technical / affective uncertainty' (Q2).
- When a participant believed that a submitted feedback passage had directly addressed an aspect of their information need (Variable Label: **[blend]**), they frequently experienced lower levels of 'technical / affective uncertainty' (Q2).
- When a participant experienced lower levels of 'conceptual / cognitive state' uncertainty (Q1), they frequently experienced lower levels of 'technical / affective uncertainty' (Q2).

The combination of Q4, **[blend]**, and Q1, provided a very reliable model of 'technical / affective uncertainty'. Interesting, Q3 was found to be a very poor predictor (Pearson correlation = -0.442). It can be seen that an increase in the amount of information that defines a domain, without addressing the information need directly, increases levels of 'conceptual / cognitive state' (Pearson correlation = 0.889, sig. at the 0.01 level).

A participant's age, gender, occupational background, online search experience, and relevance feedback experience did not significantly affect Q4, **[blend]**, and Q1. Therefore, **Hypothesis 1 (V: A, B, C, D, E)** is supported (*see* Section 10.1).

TECHNICAL / AFFECTIVE UNCERTAINTY

Model 1: Indicative or invitational mood.
Model 2: Extent an information need has been addressed + Model 1
Model 3: Conceptual / cognitive state uncertainty + Model 2

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	1.79	.699	23.267	.001
2	1.18	.183	13.990	.005
3	.71	.080	16.715	.003

TABLE 11.7: DERIVING A MODEL OF DOMAIN EXPERTISE

11.8 TASK COMPLEXITY

The classification of problem stages encountered during information seeking enables the progress of individuals to be monitored. Exploring more subtle changes in a user's problem state (e.g., monitoring completion of sub-goals) requires a different approach. Although task complexity is associated with wide range of factors, an individual's ability to pre-determine task requirements is a key factor (*see* Section 4.2.4). The feedback and answer construction process implemented by this experimental procedure is compatible with the Geneplore Model (*see* Section 9.2). The Geneplore model allows the behavioural changes associated with a users inability to pre-determine task requirements to be investigated. When an individual discovers concepts that turn out to be much more important than the ones they were contemplating at the beginning, the level of task complexity becomes higher. As task complexity increases, implications for cognitive load and behaviour might arise.

The process of retrieving existing structures from memory is analogous to selecting a passage from a document for feedback purposes. It can be assumed that the reason why a participant selects a passage is because it has a novel property, which in turn facilitates the finding of attributes, which may enable the creation of conceptual combinations that are needed for associations to be made. The number of passages highlighted by an individual *may* reflect the complexity of the task, since more novel properties may need to be found (Variable Label: [**highlight**]).

The process of association was investigated by asking a participant if they fully understood the contextual meaning of the selected passage (Variable Label: [**context**]). It can be assumed that if a participant suggested that they did not fully understand, then this revealed ambiguity. This ambiguity *might* increase task complexity. Conceptual interpretation is likely to be triggered by ambiguity, which in turn may enable synthesis.

The process of synthesis was investigated by asking participants to evaluate the utility (in terms of addressing their information need) of a selected passage (Variable Label: [**blend**]). If a participant was able to perform an evaluation of utility, this indicated that a mental blend had occurred (as the evaluation of a specific passage is performed with respect to all other passages and existing knowledge). If a passage was deemed only to be potentially useful, then this indicated that a search for meaningfulness was underway. A search for meaningfulness *might* increase task complexity. Functional inference is likely to be triggered by a search for meaningfulness, which in turn may enable transformation.

The process of transformation is represented by the way a participant transfigured a selected passage, i.e., modified the semantics by altering the feedback passage text when constructing an answer to the task. A *pe* = 2 score indicated that a passage had been integrated in some way with one or more different concepts (Variable Label: [**pe=2**]). The assignment of *pe* = 2 inferred the existence of category exemplars. The existence of category exemplars *might* increase task complexity. The process of emergence allows for contextual shifting, which in turn may result in analogical transfer.

The process of analogical transfer was investigated by observing a participant constructing an answer using previously unincorporated passages. This indicated the existence of a mental model. An incongruity within the mental model is likely to trigger hypothesis testing, which allows for categorical reduction. Incongruities could be inferred when a user was able to integrate a passage, initially identified as not being not useful for answer construction (Variable Label: [**ie=1, pe=3**]). Incongruities *might* increase task complexity.

The process of categorical reduction was investigated by asking a participant to evaluate their solution to the information problem. This evaluation process required the need for the analysis of verbal combinations. A divergence between mental and verbal models is likely to trigger a search for limitations. The task fulfilment evaluation measure was self-reported by participants (Variable Label: [**tfe**]). A low evaluation score *might* be indicative of the task complexity imposed on an individual.

The following variables, defined above, were used in an attempt to model task complexity: [**highlight**], [**context**], [**blend**], [**pe=2**], [**ie=1, pe=3**], [**tfe**]. It was decided that the overall measure of complexity should be the 'number of answer constructions' attempted (Variable Label: [**constructions**]). This was based on the assumption that the complexity of the task is associated with difficulty experienced during answer construction.

A stepwise regression analysis (*see* Table 11.8) revealed the factors that influenced the 'number of answer constructions' attempted. The results showed that task complexity was more accurately modelled when a wide range of factors were taken into consideration (standard error of the estimate = 1.29). The best individual predictor is the 'incongruities identified' measure [**ie=1, pe=3**]. Findings suggested that when an individual discovered concepts that turned out to be much more important than the ones they were contemplating at the beginning, task complexity increased.

Unsurprisingly, it was found that task complexity was reduced when a high level of contextual understanding was possessed [**context**], and when an information need was being directly addressed [**blend**] (Pearson correlation = -0.809). More interestingly, it can be seen that as participant's ability to highlight passages increased [**highlight**], task complexity was reduced. This is a counter intuitive finding. It was anticipated that task complexity would increase due to an individual encountering novel properties.

The [**pe=2**] measure was not a suitable predictor. This meant that the likelihood of category exemplars increasing task complexity could not be investigated. The [**tfe**] measure was also not a suitable predictor. The process of categorical reduction, therefore, could not be investigated.

NUMBER OF ANSWER CONSTRUCTION INTERACTIONS

Model 1: Incongruities identified.
Model 2: Contextual understanding + Model 1
Model 3: Extent an information need has been addressed + Model 2
Model 4: Number of passages highlighted + Model 3

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	2.21	.791	37.888	.000
2	1.83	.081	5.694	.041
3	1.55	.046	4.455	.068
4	1.29	.033	4.588	.069

TABLE 11.8: DERIVING A MODEL OF TASK COMPLEXITY

A participant's age, gender, occupational background, online search experience, and relevance feedback experience did not significantly affect [**highlight**], [**context**], [**blend**], [**ie=1, pe=3**]. Therefore, **Hypothesis 1 (VI: A, B, C, D, E)** is supported (*see* Section 10.1).

11.9 SUMMARY OF BEHAVIOURAL MODELS DERIVED

Five models that attempt to explain an individual's behaviour have been derived. These models were used in an attempt to predict the intrinsic and extraneous cognitive load imposed on an individual during information seeking.

- **Model 1:** Evidence was found to support the assumptions that 'conceptual / cognitive state' uncertainty and 'technical' / 'affective' uncertainty were reliable measures of **knowledge obtained from previous search interactions** (*see* Section 11.3).

- **Model 2:** Evidence was found to suggest that the *pe=2* measure could be reliably used as suitable measure of an individual's **goal-directed / exploratory behaviour**. Goal-directed behaviour could be observed when participants decided to integrate concepts during answer construction (this type of behaviour was assigned a *pe=2* score, *see* Section 11.5).

- **Model 3:** Evidence was found that if a participant was able to identify passages as not being useful, this implied that 'conceptual and cognitive state' uncertainty was low enough to allow this behaviour. If a participant was able to identify passages as being potentially useful, but was unable to integrate these within an answer construction, this perpetuated 'technical / affective' uncertainty. Results showed that an 'Initial Utility Judgement' was better explained when both 'conceptual and cognitive state' and 'technical / affective' uncertainty were considered at the same time. Although they measured separate aspects, when combined, they were likely to be more effective in assessing **overall uncertainty** (*see* Section 11.6).

- **Model 4:** Exposure to threatening, contradictory, or complex information is likely to increase levels of anxiety. A **domain expert**, however, was likely to use a schema-driven approach to information seeking. Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves without being driven by a need to reduce affective anxieties. The combination of Q4, [blend], and Q1, provided a very reliable model of 'technical / affective uncertainty'. Lower levels of 'technical / affective uncertainty' were found when: a participant suggested that they required specific information to complete a task (Q4); a participant believed that a submitted feedback passage had directly addressed an aspect of their information need [blend]; a participant experienced lower levels of 'conceptual / cognitive state' uncertainty (Q1). (*See* Section 11.7).

- **Model 5:** Results showed that **task complexity** was more accurately modelled when a wide range of factors were taken into consideration. The best individual predictor was the 'incongruities identified' measure [**ie=1, pe=3**]. Findings suggested that when an individual discovered concepts that turned out to be much more important than the ones they were contemplating at the beginning, task complexity increased. (*See* Section 11.8)

Other than the relationships reported above, no significant relationships existed amongst models. Therefore, Hypothesis 2 can be supported.

11.10 PREDICTING INTRINSIC AND EXTRANEOUS COGNITIVE LOAD

The degree of intrinsic load is defined in this thesis as the variance between usefulness judgements assigned to a document (by a participant) during and after the search process (*see* Equation 11.1). The usefulness of a document can be assessed on an individual basis, independently of other documents evaluated, and is not influenced by browsing more documents (*see* Section 11.1). A usefulness judgement is not a surrogate measure for the increased confidence experienced by a participant on completion of a task. A usefulness judgement can, therefore, be considered as a measure of situational needs. Explanations for the variance between initial and post-task usefulness evaluations were attributable to an individual's situational context.

Intrinsic cognitive load is considered as a barrier to providing an accurate usefulness judgement during the search process. Finding explanations for the variance between judgements was achieved by exploring the factors that potentially instigate behavioural change. The five models, summarised in Section 11.9, may allow the intrinsic and extraneous cognitive load imposed on an individual to be predicted.

During experimentation, the intrinsic cognitive load imposed on an individual was calculated for each individual document browsed. A comparative analysis of intrinsic cognitive load measures across a range of tasks/participants cannot be performed without normalisation. A high level of residual intrinsic cognitive load was inevitable if task performance was poor. A low level of residual intrinsic load was expected when task performance was good. A baseline measure enabled intrinsic measures of comparison to be normalised. It was decided that a self-evaluated task fulfilment score of 15 or over required no baseline as it can be anticipated that residual cognitive load was minimal. A score of 10-14 was assigned a baseline of 1.22: some residual load was anticipated (*see* Section 11.2 and Equation 11.1). A score of 6-10 was assigned a baseline of 1.38: significant residual load was anticipated. No task was assigned a score lower than six. 'Mean initial v post-task passage utility variance' (*see* Section 9.3) was used to calculate the values of the baseline measures. During the answer construction process participants who made more modifications were likely to be engaged in a more active learning / problem-solving process (*see* Chapter 12). This indicated that the cognitive load imposed on them was not high enough to prevent the modification of heuristics (*see* Section 9.2).

$$imc_j = \frac{icl}{bicl} = \frac{|iuj_j - puj_j|}{\left(\dfrac{puj_j}{tfe}\right)}$$

Where

icl	= Intrinsic Cognitive Load
$bicl$	= Baseline Intrinsic Cognitive Load
iuj_j	= Initial Usefulness Judgement assigned to document j
puj_j	= Post-task Usefulness Judgement assigned to document j
tfe	= Task Fulfilment Evaluation

EQUATION 11.1

Equation 11.2 below can be used to estimate the degree of extraneous cognitive load during the search process. The degree of extraneous load is the variance between the evaluation of usefulness of a feedback passage (submitted by a searcher) during and after the search process. The initial evaluation of passage usefulness can be performed on an independent basis, without being influenced by browsing more documents (*see* Section 11.1). The mean evaluative score did not significantly improve or worsen as more documents were browsed. A post-task passage usefulness score was not significantly related to the initial evaluation of passage usefulness (*see* Section 9.3). Passage usefulness assignments are not surrogate measures for the increased levels of knowledge possessed by a participant on completion of a task. A passage usefulness judgement can, therefore, be considered as a measure of situational needs.

Extraneous cognitive load is considered as a barrier to providing accurate feedback judgements during the search process. When an initial usefulness evaluation (*ie*) was made for passage *pj*, and compared to a post-search usefulness evaluation (*pe*) for passage *pj*, extraneous load was calculated. Finding explanations for the variance between judgements was achieved by exploring the factors that instigate behavioural change. The five models, summarised in Section 11.9, may allow the intrinsic and extraneous cognitive load imposed on an individual to be predicted.

$$ecl = \left| ie_{pj} - pe_{pj} \right|$$

Where

ecl = Extraneous Cognitive Load
ie_{pj} = Initial evaluation of passage usefulness for pj
pe_{pj} = Post-search evaluation of passage usefulness for pj

EQUATION 11.2

A range of behavioural factors can *potentially* affect the generation of feedback judgements by inducing cognitive load. The development of the five models, summarised in Section 11.9, enabled prospective 'cognitive load model' contributors to be evaluated using a stepwise multiple regression analysis. The intrinsic (*imc*) and extraneous (*emc*) multiple regression analyses are shown in Equation 11.3.

Predicted $imc\&emc = b_1 KnowledgeObtainedFromPreviousInteractions + b_2 GoalDirectedBehaviour + b_3 OverallUncertainty + b_4 DomainExpertise + b_5 TaskComplexity + b_0$

Where

b_n = testing regression coefficients determines whether potential contributors to the model have a non-zero relationship to the intrinsic or extraneous measure of comparison, i.e., this is a measure of the steepness of the best-fit lines.

b_0 = regression constant, i.e., the intercept.

EQUATION 11.3

The aim was to identify a small combination of IVs that account for the majority of variance in the DV. The use of intrinsic and extraneous measures of comparison enabled qualitative information, derived from users' explanations of their situation and decisions, to enable causal, quantitative explanations. Results showed that the prediction of intrinsic and extraneous cognitive load is possible (*see* Table 11.9). It was somewhat surprising that for both intrinsic and extraneous load, only a single behavioural factor was found to be a suitable predictor. However, these behavioural factors were different ones. This suggested that both intrinsic and extraneous elements must be represented within a model of cognitive load. **Hypothesis 3 (II, III, IV, V, VI) and Hypothesis 4 (I, II, III, IV, V)** were supported as no significant relationships were found (*see* Section 10.1). However, **Hypothesis 3 (I) and Hypothesis 4 (VI)** were not supported. These findings are discussed in Chapter 12.

- **'Knowledge Obtained From Previous Search Interactions' is an excellent predictor of 'Intrinsic Cognitive Load'**
  (standard error of the estimate = 1.25)
- **'Task Complexity' is an excellent predictor of 'Extraneous Cognitive Load'**
  (standard error of the estimate = 1.36)

INTRINSIC COGNITIVE LOAD

KNOWLEDGE OBTAINED FROM PREVIOUS SEARCH INTERACTIONS

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	1.25	.740	28.437	.001

EXTRANEOUS COGNITIVE LOAD

TASK COMPLEXITY

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	1.36	.754	30.638	.001

TABLE 11.9: INTRINSIC AND EXTRANEOUS MODELS OF COGNITIVE LOAD

12.0 INTRODUCTION

Results detailed in Chapter 11 imply that individuals are not consistent in applying the same behavioural style for every task they undertake. It is theorised that new search heuristics are generated when an individual is unable to assimilate information during the search process. Cognitive load is caused by the uncertainty that arises when an individual has to apply new heuristics, and when a mismatch in perceptions and actual experience causes existing heuristics to be abandoned. Predicting the utility of feedback judgements using Cognitive Load Theory is possible. This Chapter will investigate the six conjectures proposed in Section 10.1.

12.1 COGNITIVE LOAD AND INTERACTION WITH INFORMATION OBJECTS

Cognitive flexibility theory is a constructivist theory that was specifically formulated to investigate the use of interactive technology, e.g., hypertext. Spiro and Jehng (1990) were the initial pioneers of the theory. Cognitive flexibility theory is largely concerned with the transfer of knowledge and skills beyond their initial learning situation. The theory asserts that effective learning is context-dependent and learners must be given the opportunity to develop and integrate their own representations of information in order to properly learn/understand.

When using an IR system, a user needs to be able to spontaneously restructure knowledge as an information need develops. The dynamic nature of an information need means that if knowledge can be represented along multiple rather than single conceptual dimensions then the assembly of new schemata is aided. Therefore, an individual's investigation of an information need is likely to require a combination of exploratory and goal-directed behaviour. Goal-directed behaviour requires an individual to develop their own representations of information.

Evidence was found to suggest that the *pe=2* measure could be reliably used as suitable measure of an individual's goal-directed / exploratory behaviour. Goal-directed behaviour could be observed when participants decided to integrate concepts during answer construction.

Extraneous cognitive load influences an individual's ability to interact with information objects (*see* Chapter 7). However, no evidence was found that suggested goal-directed / exploratory behaviour could be attributed to the extraneous cognitive load imposed on an individual (*see* Section 11.10). The ability to be cognitively flexible did not seem to significantly affect the amount of cognitive load. Without further investigation it is difficult to suggest an explanation for this.

It has been discovered that task complexity (*see* Section 11.8) was responsible for extraneous cognitive load, and consequently, could be partially responsible for the way individuals interacted with information objects. Individuals who were able to spontaneously restructure knowledge in response to situational demands experienced less extraneous cognitive load. When an individual discovered concepts that turned out to be much more important than the ones they were contemplating at the beginning, task complexity increased.

Pask (1976a) discovered that individuals could be categorised as being 'holists' or 'serialists'. Holists adopt a global approach by examining relationships between a range of potentially relevant topics and building an overview. Serialists prefer to concentrate on one topic at a time. Consequently, serialists obtain an overview at a much later stage. Interestingly, Pask also found that serialists were most affected by uncertainty.

Participant who experienced high levels of extraneous cognitive load submitted a fewer number of relevance feedback passages and found it difficult to pre-determine task requirements. This could be observed when a participant identified a passage as initially not being useful but then used it later within the answer construction process. This behaviour indicated a reliance on serialist information processing, and the lack of a developed task overview. It was found that participants who in some way attempted to incorporate a passage into an answer construction experienced a lower level of extraneous cognitive load. This behaviour was indicative of an attempt to develop a task overview. This is a typical holistic strategy.

Pask (1976b) associated holist and serialist approaches to components of understanding. The holist strategy was related to 'comprehension learning', while the serialist strategy was related to 'operational learning'. These differences have implications for interaction. Entwistle (1981) suggested that a 'deep approach' involves evaluating the extent to which conclusions can be supported and integrated within an individual's knowledge, while a 'surface approach' is limited to learning specific facts that are often disconnected.

Witkin et al's (1977) well-established concept of field dependent/independent cognitive styles seems to be compatible with both Pask's and Entwistle's findings. Field-independent individuals incorporate new information into a framework. Field-dependent individuals tend to rely on and preserve the structure of information. Consequently, field-independent individuals are better at organising information, which is analogous to the holist/deep approach.

Exposure to threatening, contradictory, or complex information is likely to increase levels of anxiety. A domain expert, however, is more likely to use a schema-driven 'field-independent' approach to information seeking. Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves without being driven by a need to reduce affective anxieties (*See* Section 11.7). However, no evidence was found that suggested that a domain expert experienced lower levels of cognitive load (*see* Section 11.10).

Expertise did not seem to significantly affect the amount of cognitive load. A possible explanation was that adopting a 'schema-driven field-independent approach' does not necessarily mean that a holistic approach to the assimilation of information can always be undertaken. Individuals were not consistent in applying the same behavioural style for every task they undertook. It is theorised that new search heuristics are generated when an individual is unable to assimilate information during the search process. Heuristics are investigated in Section 12.5.

12.2 COGNITIVE LOAD DURING RELEVANCE FEEDBACK

Cognitive load experienced during relevance feedback is the intrinsic and extraneous difficulty associated with providing useful / accurate feedback judgements during information seeking. Barry (1994) discussed the potential of using clues to identify characteristics of documents that could be used by an IR system to retrieve documents (i.e., going beyond topicality). However, the interpretation of perceptual clues are dependent on structures of goals held in memory and expertise, i.e., a personal perspective (Chase and Simon, 1973). Importantly, identifying the characteristics of perceptual clues, where the discovery of novelty is an essential element, cannot be achieved from post-hoc generalisations. The lack of personal perspective may be the result of cognitive load imposed on an individual. More useful / accurate feedback judgements will be made if both an individual's intrinsic cognitive load and extraneous cognitive load is low. We currently have little knowledge of how uncertainty or different cognitive styles affect information behaviour. It is currently impossible to speculate how uncertainty or adopting a particular cognitive style triggers transfigurations in schemata.

Kuhlthau's (1993a; 1993b) 'uncertainty principle' has served as a valuable insight into the nature of uncertainty. Experimentation attempted to explore Kuhlthau's ideas by attempting to incorporate affective aspects of uncertainty within a model of information behaviour. The possibility that manifestations of uncertainty are symptomatic of limitations to cognitive capabilities has been considered. The ability to assimilate new information is burdened by cognitive load. Kelly's (1963) personal construct theory is an attempt to explain the affective experiences of individuals. Kelly suggested that when new information is assimilated, confusion increases as inconsistencies and incompatibilities are found. If the information becomes too threatening, then it is discarded. Kelly believed that the user's mood (invitational or indicative) dictates the ease with which information can be assimilated. No evidence was found that suggested goal-directed / exploratory behaviour could be attributed to the cognitive load imposed on an individual (*see* Section 11.10). However, a user's indicative mood was correlated with lower levels of 'technical / affective uncertainty' (*see* Section 11.10).

Dervin (1983) suggested information seeking could be viewed as a process of sense-making. An individual is actively involved in finding meaning that fits in with what they already know within a personal framework of reference. This makes the process of identifying novel concepts during the process of providing feedback difficult.

Kuhlthau (1991) argued that a model representing a user's sense-making should incorporate: physical, actual actions taken; affective, feelings experienced; and cognitive, thoughts concerning both process and content. Since people have a limited capacity of assimilating new information, they purposefully construct meaning by selectively attending to that which connects with what they already know. Kuhlthau speculated that the mismatch of user perception and actual experience might increase confusion and anxiety. The representation of the intrinsic cognitive load concept enables Kuhlthau's speculation to be tested. Intrinsic cognitive load is a measure of a mismatch between perceptions (an initial utility judgement) and actual experience (post-task utility judgement) (*see* Section 9.1).

Evidence was found to support the assumptions that 'conceptual / cognitive state' uncertainty and 'technical' / 'affective' uncertainty were reliable measures of knowledge obtained from previous search interactions (*see* Section 11.3). It has been discovered that knowledge obtained from previous search interactions was responsible for intrinsic cognitive load, and consequently, could be partially responsible for the way individuals assimilate information. Individuals who were able to manage uncertainty in response to situational demands experienced less intrinsic cognitive load (*see* Section 11.10).

12.3 INTRINSIC AND EXTRANEOUS COGNITIVE LOAD

Intrinsic load is the individual-specific difficulty associated with a retrieval task (dependent on user expertise and task complexity). Extraneous load is the individual-specific difficulty associated with identifying pertinent information (dependent on learner activity).

Intrinsic factors can be conceptualised as components of an individual's 'anomalous state of knowledge' (Belkin, 1977), or 'state of uncertainty' (Ingwersen, 1992). Essentially, intrinsic factors are components of an individual's knowledge state, which is comprised of schemata (Cooper, 1990), which are accessed using heuristics (Simon, 1969). The state of an individual's knowledge governs the amount of cognitive load imposed on the individual. Sophisticated knowledge states can be held more easily in working memory. Less sophisticated knowledge states cannot be held entirely within working memory so a higher cognitive load is imposed. These principles are identical to the ones proposed by all researchers who apply Cognitive Load Theory to their various research fields (*see* Section 6.2). No new assumptions about an individual's cognitive architecture have been made. Knowledge obtained from previous search interactions is responsible for intrinsic cognitive load, and consequently, could be partially responsible for the way individuals assimilate information.

Extraneous factors can be conceptualised as components of an individuals 'work space' (Ingwersen, 1992) or 'naming and projected structure stage' (Fauconnier, 1997). Extraneous load is governed by the suitability of the heuristics chosen for addressing an individual's information need at a specific point in time. More suitable heuristics enable information to be processed in a way that is more compatible with existing knowledge schemata. Task complexity (*see* Section 11.8) is responsible for extraneous cognitive load, and consequently could be partially responsible for the way individuals interact with information objects. Individuals who were able to spontaneously restructure knowledge in response to situational demands experienced less extraneous cognitive load (*see* Section 11.10). Selecting suitable heuristics that can, in some way anticipate holistic task needs, reduces task complexity and extraneous cognitive load.

12.4 UNIVERSAL CAUSAL AND ASSESSMENT FACTORS

Both causal (intrinsic) and assessment (extraneous) factors affect cognitive load. Causal factors are the characteristics of an individual (e.g., cognitive abilities, and domain knowledge), while assessment factors are characteristics of the mental effort needed and the mental load imposed. Previous studies[34], which used Cognitive Load Theory, reported similar conclusions that are possibly universally applicable to many domains. The following aspects are known to impose a heavy cognitive load: inappropriate problem solving methods; integrating several sources of information; and unnecessarily repetitive information. The Cognitive Load Theory proposed within this thesis drew similar conclusions (*see* Section 12.3). Knowledge obtained from previous interactions can cause problems when assimilating (integrating) several sources of information. Task complexity can cause inappropriate heuristics to be selected (problem solving methods). A universal aspect not investigated was that of 'unnecessarily repetitive information'. Documents presented to experiment participants contained conceptually different information. Future research could investigate this aspect by presenting documents that conceptually overlap.

The natural progression of the information seeking process ensures that behaviour progresses through a number of characteristic stages. A number of information seeking behaviour models exist (*see* Chapter 4). The majority have attempted to represent search behaviour by modelling a single process. Wilson (1999) believed that a unifying framework is needed to express the structure and the processes involved in successive searching. This is an aspect that previous models have not addressed. Wilson identified that a key part of the problem solving process is that of 'uncertainty reduction'. Schutz (1967) believed that stored experiences contain non-ostensive characteristics that can be used to perceive new experiences. From this, Wilson hypothesised that when uncertainty fails to be resolved, this may result in a feedback loop to a previous stage in the problem solving process. Increasing the certainty associated with a previous stage may reveal new characteristics that can be used to perceive new experiences.

[34] Ayres, 1993; Paas, 1992; Van Merrienboer and Krammer, 1987; Cooper and Sweller, 1987; Sweller, Chandler, Tierney and Cooper, 1990.

Wilson also believed that during goal-seeking behaviour, an individual moves from uncertainty to increasing certainty and that there are stages in the problem resolution process that are identifiable and recognisable to an individual (*see* Figure 12.1).

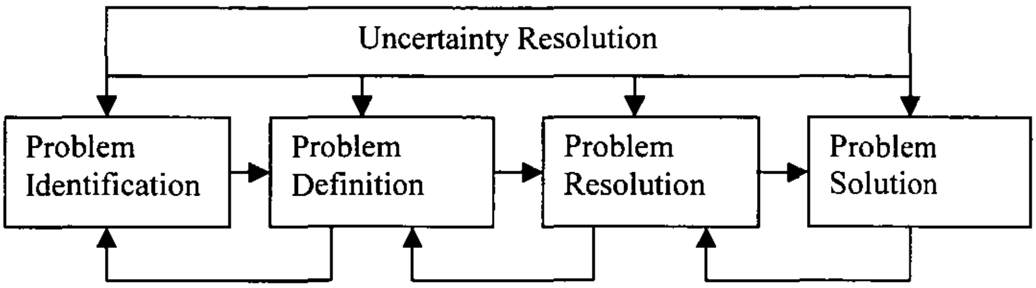


FIGURE 12.1: UNCERTAINTY RESOLUTION (Source: Wilson, 1999)

Although Wilson's problem resolution chain model was a worthwhile attempt at enabling the unification of information seeking models, it is not suitable for modelling the feedback process investigated within this thesis. Wilson's reliance on Schutz's 'uncertainty' conjecture should be called into question. A 'global uncertainty' measure is assumed to trigger a feedback loop.

Evidence was found (*see* Section 11.6) that if an experiment participant was able to identify passages as not being useful, this implied that 'conceptual and cognitive state' uncertainty was low enough to allow this behaviour. If a participant was able to identify passages as being potentially useful, but was unable to integrate these within an answer construction, this perpetuated 'technical / affective' uncertainty. Results show that an 'Initial Utility Judgement' can be better explained when both 'conceptual and cognitive state' and 'technical / affective' uncertainty are considered at the same time. Although they measure separate aspects, when combined, they are likely to be more effective in assessing overall uncertainty. However, it was discovered that a measure of overall uncertainty was not a suitable predictor for the level of intrinsic or extraneous cognitive load imposed on an individual. This finding suggested that the use of a global uncertainty measure is unsuitable for modelling feedback utility. An alternative approach is required. This approach might also be suited to modelling a unifying problem solving structure and the processes involved in successive searching (*see* Figure 12.2).

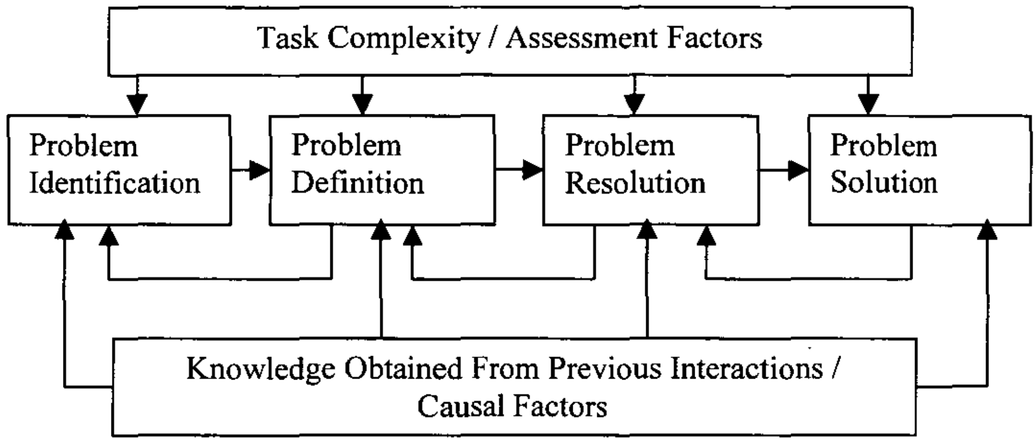


FIGURE 12.2: CAUSAL AND ASSESSMENT FACTORS

Figure 12.2 shows that a feedback loop is more likely to be triggered by causal and assessment factors rather than global uncertainty. For relevance feedback purposes knowledge obtained from previous interactions is a powerful causal factor; and task complexity is a powerful assessment factor. If either of these factors are found to impede the resolution of the next problem solving stage, this might trigger a feedback loop to a previous stage. It must be emphasised that different causal and assessment factors need to be investigated to provide a generalizable model. The flexibility of using Cognitive Load Theory may allow new causal and assessment factors to be identified for different types of information seeking behaviour, i.e., not just for relevance feedback.

12.5 HEURISTICS

Intrinsic factors are components of an individual's knowledge state, which is comprised of schemata, which are accessed using heuristics. Extraneous load is governed by the suitability of an individual's heuristics at processing information objects. More suitable heuristics enable information to be processed in a way which is more compatible with existing knowledge schemata. An argument, detailed in Section 12.1, proposed that individuals are not consistent in applying the same processing style for every task they undertake. It is theorised that new search heuristics are generated when an individual is unable to assimilate information during the search process.

Theories of field dependent and field independent cognitive styles have a number of implications for information seeking behaviour. Witkin and Goodenough (1981) identified a number of other cognitive styles that have an impact on learning: Differences in the extent and intensity of attention; Differences in remembering distinct memories versus the tendency to merge similar events; Speed and adequacy with which alternative hypotheses are formed and responses made. Although it is possible to observe some types of cognitive behaviour during information seeking, most cognitive styles cannot be observed directly. Categorising participants as possessing a particular cognitive style by performing psychometric testing is often the only method. It is proposed that individuals are not consistent in applying the same processing style for every task they undertake. Categorising an individual as a holist or serialist is not always a valid approach (*see* Section 12.1).

Pask (1976b) and Witkin et al. (1977) both found that learning outcomes were positively affected by presenting the same information in a way that matched an individuals learning style. Daniels (1986) claimed that when developing user models, crude over-generalisations might arise when the system is forced to match a user with a preconceived behavioural template. Individuals undertake information seeking to facilitate understanding. Pollock, Chandler, and Sweller (2002) found that artificially reducing the element interactivity of complex information allows elements to be processed serially.

Results from Pollock et al's study indicated that for complex information processing, information is often better understood through isolating interacting elements. This finding is of considerable importance to the study of information behaviour. It may help explain one of the reasons why individuals undertake different information seeking (or browsing) behaviour even though they can be categorised, by psychometric testing, as possessing a preferred cognitive style. During the answer construction process, participants who made more modifications were likely to be engaged in a more active learning / problem-solving process. This indicated that the cognitive load imposed on them was not high enough to prevent the modification of heuristics. It is theorised that new search heuristics are generated when an individual is unable to assimilate information during the search process. These heuristics may be used to enable understanding by isolating interacting elements.

12.6 PREDICTION

Cognitive Load Theory applied to the domain of problem solving provides a pragmatic insight into how information can be presented in a way that circumvents cognitive limitations. When Cognitive Load Theory is applied to the domain of learning, fewer pragmatic insights are likely as the interpretation of meaning affects understanding in ways that cannot be modelled by Cognitive Load Theory. However, Cognitive Load Theory can be a useful framework for the exploration of behavioural processes during learning.

The development of a relevance feedback system that attempts to predict usefulness is problematic. Such a system would need a model of situational usefulness that can be inferred from the user. Passage relevance feedback is most likely to allow the context of a situational need to be inferred, as selection of individual terms does not reveal context. The model of Cognitive Load developed within this thesis enables the utility of feedback to be predicted during the information seeking process. Feedback has been shown to be more accurate when an individual is able to manage the cognitive load imposed by identifying 'usefulness clues'. These clues trigger new heuristics that may be more appropriate for addressing an individual's information need.

Bruner (1986, 1990) believed that IR should ideally be a structured process so that it can facilitate an individual's understanding. Studying information behaviour allows an understanding of the experiences and contexts that enable an individual to successfully retrieve information. Figure 12.3 shows that intrinsic and extraneous cognitive load has the potential to affects an individual's heuristics (search strategies) in different ways. It is theorised that the intrinsic cognitive load imposed on an individual limits the type of heuristics deemed suitable by that individual for addressing their information need. Depending on an individual's capabilities, the most suitable repertoire of heuristics will be selected by them. This selection is based on the knowledge they have obtained from previous interactions (*see* Section 11.3). Individuals who were able to manage uncertainty in response to situational demands experienced less intrinsic cognitive load.

It is theorised that the extraneous cognitive load imposed affects the way in which heuristics are modified. Modification of heuristics is essential to enable information objects to be processed in a way that is compatible with an individual's existing knowledge schemata. Task complexity (*see* Section 11.8) is responsible for extraneous cognitive load, and consequently, could be partially responsible for the way individuals interact with information objects. When an individual discovers concepts that turn out to be much more important than the ones they were contemplating at the beginning, task complexity increases. Modifying heuristics that can, in some way anticipate holistic task needs, reduces task complexity and extraneous cognitive load.

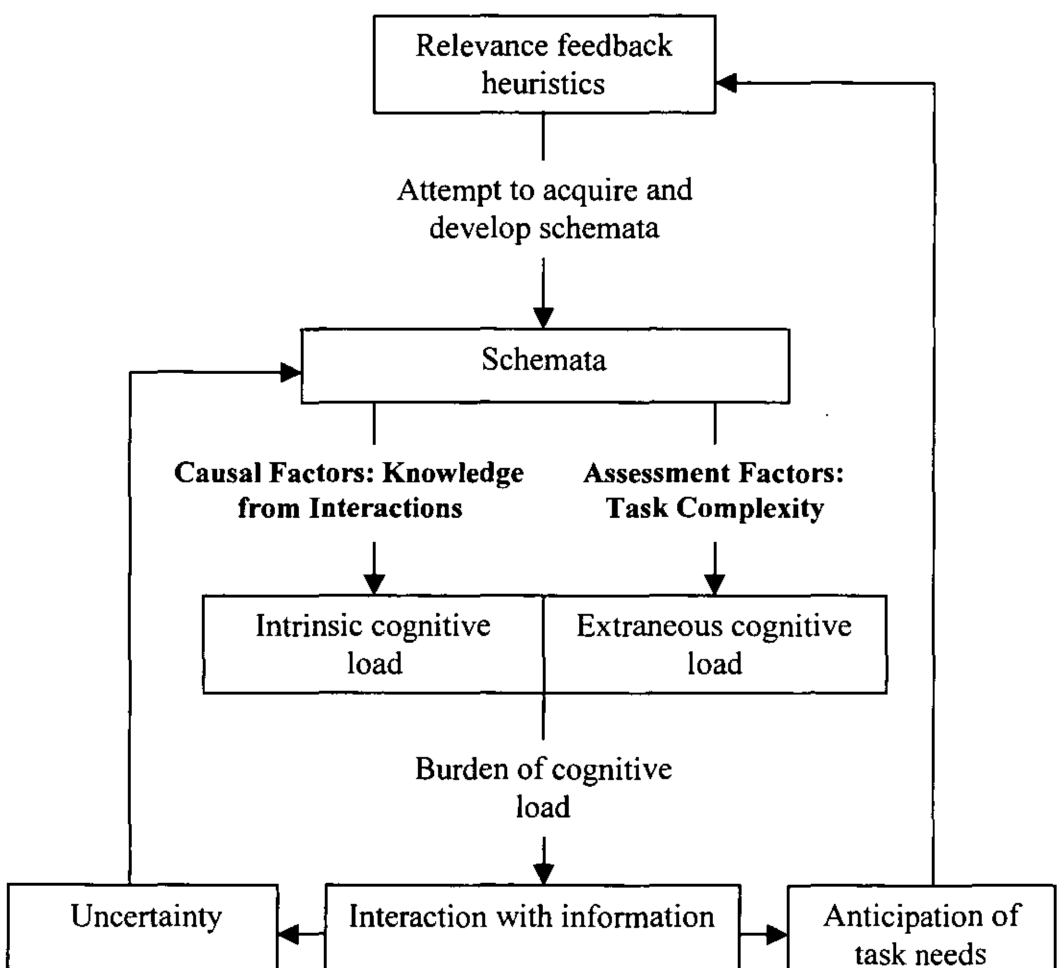


FIGURE 12.3: COGNITIVE LOAD THEORY FOR RELEVANCE FEEDBACK

It is proposed that a feedback mechanism could allow the burden of cognitive load to be modelled. This would require the monitoring of causal and assessment factors during information seeking. The aim is to aid the discovery of new heuristics, which may be more suitable for addressing an individual's information need.

An answer construction process, similar to the one detailed in Section 10.3.3, might allow usefulness incongruities to be identified. An individual's usefulness judgements made during an earlier stage of the search process were generated using less developed knowledge schemata. The analysis of an answer construction process may reveal causal and assessment factors which previously resulted in a high cognitive load being imposed.

13.0 INTRODUCTION

Outside the laboratory environment, user-based relevance feedback systems are rarely utilised when implemented. An evaluation procedure has been developed that allows an insight into why such feedback is often not effective. An individual's cognitive limitations affect information behaviour and dictate the way in which new information is assimilated. Evaluation requires the use of Cognitive Load Theory to reveal the context in which feedback judgements are made. Ultimately, this will allow the utility of relevance feedback submitted by a user to be predicted. The usefulness of being able to predict the utility of feedback has pragmatic implications for developing more intuitive IR systems. Interaction is to some extent driven by cognitive load. This thesis has shown that it is possible to identify behavioural manifestations that are symptomatic of cognitive load. By identifying judgement incongruities, IR systems can be designed to prompt an individual to re-evaluate previous judgements that have been affected by high levels of cognitive load. This allows for more accurate feedback judgements that increase the likelihood of successful IR.

Behavioural explanations elicited from individuals are burdened by cognitive load. The intention was not to set up a situation where experiment participants were required to explain their information interactions. The model of user feedback interaction developed within this thesis was derived from a quantitative investigation of ideas developed through qualitative research. When examining cognitive processes, it is impossible to discover when a sufficient number of properties have been found. This is because humans do not perform exhaustive searches of a problem space, even if the problem is simple enough to make it feasible. Humans use heuristics instead. Heuristics are complex cognitive structures that cannot be predicted. The basic dilemma with modelling interaction is reconciling the need for prediction with the seemingly individual-specific effect of information. Interaction is guided by heuristics and not by logical analysis or deduction. Heuristics impose assumptions that are used to address a problem in a way that is compatible with an individual's knowledge schemata. The generation and application of heuristics is influenced by the cognitive load imposed on an individual.

The Cognitive Load Theory developed within this thesis has both explanatory and predictive power. Demonstrating how the burden of cognitive load can be assessed within the context of providing relevancy judgements has been achieved through an understanding of learning structures. Limited mental resources mean that only a few elements of information may be attended to at any given time. What constitutes an element of information depends on the schemata held by a person. A single element consisting of a single schema for an expert may be several elements consisting of several sub-schemata for a novice. This research shows that individuals are not consistent in applying the same processing style. New search heuristics are generated when an individual is unable to assimilate information during the search process. Cognitive Load Theory cannot explain the nature of the heuristics an individual applies during information seeking. However, by modelling changes in cognitive load, identification of changes in heuristics is possible. Changes in heuristics can explain why an individual has adopted a new behavioural approach. Cognitive load is caused by the uncertainty that arises when an individual has to apply new heuristics, and when a mismatch in perceptions and actual experience causes existing heuristics to be abandoned.

These findings imply that the further development of Cognitive Load Theory could provide information science with a framework that enables the examination of how an individual's cognitive architecture affects interaction with information objects. Focussing on an individual's interaction with information objects enables a connection to be made between cognitive/behavioural assumptions and real situations. It is anticipated that the explanatory and predictive power of the proposed model will improve as more studies of information behaviour are incorporated. The novel method adopted by this thesis enables a range of behavioural factors to be used to identify symptoms of cognitive load. As our understanding of information behaviour develops, the flexibility of the experimental design allows new phenomenon to be considered. Better predictors of the cognitive load imposed on an individual may well be found. This research may be the first step to providing information science with a method that enables qualitative research findings to be quantitatively evaluated, and better incorporated within the development of IR systems capable of adapting to a user's cognitive limitations.

13.1 LIMITATIONS

Investigating why feedback is a problematic process involved the inference of situational usefulness from a participant during the process of generating a judgement. Therefore, exploring why users are often reluctant to provide feedback involved experimentation that enforced the generation of feedback judgements. Enforcing users to provide feedback introduces an artificial aspect. Successfully completing a task required a written answer to be constructed. Although it is not unusual for individuals to make notes during information seeking, the answer construction processes required participants to incorporate the semantics of as many feedback passages as possible. This procedure was essential as it enabled the situational usefulness of feedback passages to be evaluated during the information seeking process.

Requiring an individual to provide a written answer using feedback passages reveals manifestations of hidden, highly abstract, cognitive constructions that cannot be otherwise observed. Mental spaces are the domains that discourse builds up to provide a cognitive substrate for reasoning and interfacing with the world. The answer constructing process enables participant feedback decisions to be continually evaluated during information seeking. A feedback passage incorporated at some stage of construction contains various grammatical devices such as descriptions regarding what new conceptual spaces are being set up, clues to which space is currently in focus, descriptions that identify existing elements, and descriptions that introduce new elements into spaces. An analysis of an answer construction resulted in an improved understanding of how providing feedback is linked to an individuals understanding of their information situation. However the artificial nature of requiring individuals to provide feedback in order to complete a task is a limitation to the validity of the experimental design, as individuals do not interact with real IR systems in the same way. This research has provided a valuable insight into how cognitive limitations affects the utility of feedback in an artificial environment. However, without further experimentation, it cannot be claimed that the results obtained are generalizable to actual IR domains. However, this research has provided evidence that cognitive load invariably affects the feedback process and must be considered as an important factor when evaluating and developing actual IR systems.

Within this research, the use of simulated retrieval tasks is another artificial aspect. Although participants were free to interpret task requirements in their own way, participants performed tasks as 'information intermediaries', i.e, the initial ASK or 'cognitive breakdown situation' was simulated. Simulated tasks were used as a means to induce the likelihood of cognitive load. Tasks were complex and required the use of multiple sources of information. The conceptual relations between sources were often not immediately apparent. This increased the possibility of 'information overload' occurring. In an actual information environment, an individual's ASK may not require complex information to be evaluated and hence might not induce a heavy cognitive load. The experimental design aimed to artificially impose cognitive load in order to magnify associated effects, making them more easily observable. Now that a model enabling the prediction of feedback utility has been developed, the next stage is to expand the scope of the model further using actual IR systems and real ASKs.

Identifying the role heuristics play during information seeking was achieved by examining the evolution of the cognitive load concept. Cognitive Load Theory is an established concept within the field of educational psychology. However, although discursive learning environments have been studied, no previous attempts have been made to use the theory to investigate information seeking behaviour. Within information science, Cognitive Load Theory enables many behavioural processes to be incorporated within a framework that provides explanations for changes in an individual's information behaviour.

This thesis has attempted to apply Cognitive Load Theory in a way that uses previous research identified by information scientists as affecting information behaviour. Consequently, an artificial experimental design was required to elicit a range of factors during information seeking: knowledge from previous interactions; exploratory/goal-directed behaviour; uncertainty, domain knowledge, and task complexity. The development of a model of cognitive load required exploratory experimentation. The nature of this experimentation meant that participants were often required to perform procedures that were somewhat convoluted.

A range of behavioural variables were elicited as this enabled a number of factors to be used as exploratory predictors of cognitive load. The validity of each behavioural factor could only be verified after experimentation. Participants were required to provide a response to the same set of questions after every interaction with a different information object. This process was repetitive and might have distracted participants from the task being addressed. The exploratory nature of experimentation may have caused motivational problems for some participants, but the vast majority remained enthusiastic or patient. An improved understanding of the type of behavioural variables that need to be elicited now exists. Future experimentation can eliminate questions that are surplus to requirements.

The limitations associated with the artificiality of the experimental design have been difficult to avoid. Understanding how cognitive load affects feedback judgements has required the development of novel procedures that explore aspects that have never been considered by information scientists. Although some behavioural patterns can be observed during information seeking, many factors require participants to self-report variables in an artificial way. An increased understanding of observable behavioural patterns may allow self-reported variables to be inferred from behavioural patterns in the future.

Demonstrating how an individual's cognitive architecture imposes limitations during feedback has been achieved by developing a new theoretical postulate. This postulate has been experimentally tested but is by no means a comprehensive model of the feedback process. The model has only identified what appeared to be the main contributors of cognitive load imposed on an individual. Other contributors no doubt exist. The proposed model is the first step towards the development of Cognitive Load Theory within information science rather than a fully formulated conception.

13.2 MAIN CONCLUSIONS

Aim One (*see* Section 1.5) was to explore why users are often reluctant to provide relevance feedback. Firstly, existing research findings were reviewed in an attempt to understand the problem. It was found that little research had been performed on the development of methods that capture relevancy judgments from a user in different ways. Exploratory Study 1 demonstrated that passage relevance feedback, a spontaneous approach, could improve the effectiveness of an automatic query expansion. However, it was concluded that evaluating the usability, associated with feedback capturing mechanisms, could not be performed using the metrics developed for Exploratory Study 1, as individual subjective differences remain hidden. For Aim One to be fulfilled, a method for evaluating relevance judgements assigned to information objects must allow judges to signal their cognitive capability during evaluation. Before a method could be developed, Aim 2 had to be fulfilled.

Aim Two (*see* Section 1.5) was to identify the role heuristics play during problem solving by examining the evolution of the cognitive load concept. Existing research findings were reviewed in an attempt to understand how heuristics and Cognitive Load Theory affect problem solving. It was found that the concept of 'judgement accuracy' has not been previously researched. A user may well be unable to communicate their cognitive state to a system. A user's perceived mental model may not be an accurate reflection of their information need situation. Individuals purposefully construct meaning by selectively attending to that which connects with what they already know.

As knowledge states shift to more clearly focussed thoughts, a parallel shift occurs in feelings of increased confidence. Commonly experienced phases or stages provide an abstract view of the search process. They do not increase our understanding of how an individual progresses from one search stage to another. This thesis has avoided making the assumption that uncertainty is a suitable measure of an individual's ability to process information. The possibility that manifestations of uncertainty are symptomatic of limitations to cognitive capabilities needs to be considered. The flexibility of Cognitive Load Theory can be used to provide an improved understanding of behaviour that integrates both cognitive and affective dimensions.

Judgements are provided at a specific point in time and will fluctuate as an information need develops. The cognitive load imposed on an individual is also dynamic. Exploratory Study 2 found that initial and post-task usefulness judgements assigned to documents could be assessed on an individual basis, independently of other documents. No confounding variables were found that could be associated with the idea that initial and post-task judgements may improve or worsen as more documents are presented. Furthermore, it was found that an initial judgement does not significantly influence a post-task judgement.

Poorly developed knowledge schemata cannot be fully integrated within working memory. Exploratory Study 2 set simulated task difficulty at a level that induced cognitive load. The majority of participants found that making notes helped them to manage tasks. Elements (separate ideas 'talked about') were included within notes because they could not be incorporated fully within existing knowledge schemata. Evidence has been found to demonstrate that cognitive load is a measurable phenomenon. The number of elements incorporated within notes was significantly positively correlated with the degree of variance between initial and post task judgements.

Exploratory Study 3 found that participants were able to self-report changes in relevance feedback behaviour. Participants recognised that different approaches are needed depending on the scenario imposed by an information need. Exploratory Study 3 helped to identify behavioural variables that could *potentially* be incorporated within a cognitive load model of feedback.

The mechanisms developed by this thesis used to estimate the amount of cognitive load imposed on an individual, and to identify the factors causing cognitive load, do not attempt to measure changes in an individual's knowledge directly. This is an unsustainable research goal. Instead, the state of an individual's knowledge state was ascertained by exploring factors that were identified by existing qualitative research as being both measurable and able to provide an insight into an individual's information behaviour.

Before it is possible to demonstrate if and how an individual's cognitive architecture imposes limitations during relevance feedback (Aim 3), and to develop a mechanism allowing the epistemic context in which judgements occur to be revealed (Aim 1), conjectures on how cognitive load affects information behaviour were hypothesised:

- Cognitive load experienced during relevance feedback is the intrinsic and extraneous difficulty associated with providing useful/accurate feedback judgements during information seeking. More useful/accurate feedback judgements will be made if both an individual's intrinsic cognitive load and extraneous cognitive load is low.

- Intrinsic load is the individual-specific difficulty associated with a retrieval task (dependent on user expertise and task complexity). Extraneous load is the individual-specific difficulty associated with identifying pertinent information (dependent on learner activity).

- Both causal (intrinsic) and assessment (extraneous) factors affect cognitive load. Causal factors are the characteristics of an individual (e.g., cognitive abilities, and domain knowledge), while assessment factors are characteristics of the mental effort needed and the mental load imposed. Mental load is the portion of cognitive load that is imposed exclusively by the task and environmental demands.

- Intrinsic factors are components of an individual's knowledge state, which is comprised of schemata, which are accessed using heuristics. Extraneous load is governed by the suitability of an individual's heuristics at processing information objects. More suitable heuristics enable information to be processed in a way which is more compatible with existing knowledge schemata.

The above conjectures were developed by identifying the role heuristics play during problem solving, by examining the evolution of the cognitive load concept, and by applying them to the domain of providing feedback judgements (Aim 2).

13.2.1 MAIN EXPERIMENTAL PROCEDURE

Evidence was found to support the assumptions that 'conceptual/cognitive state' uncertainty and 'technical/affective' uncertainty were reliable measures of knowledge obtained from previous search interactions. Goal-directed behaviour could be observed when participants decided to integrate concepts during answer construction. If a participant was able to identify concepts as not being useful, this implied that 'conceptual/cognitive state' uncertainty was low enough to allow this behaviour. If a participant was able to identify passages as being potentially useful, but was unable to integrate these within an answer construction, this perpetuated 'technical/affective' uncertainty.

Exposure to threatening, contradictory, or complex information is likely to increase levels of anxiety. A domain expert, however, is likely to use a schema-driven approach to information seeking. Such an approach enables an expert to proceed because a well-developed schema encodes a series of problem states and associated moves without being driven by a need to reduce affective anxieties. Lower levels of 'technical/affective uncertainty' were found when: a participant suggested that they required specific information to complete a task; a participant believed that a submitted feedback passage had directly addressed an aspect of their information need; a participant experienced lower levels of 'conceptual/cognitive state' uncertainty.

Task complexity can be more accurately modelled when a wide range of factors are taken into consideration. When an individual discovered concepts that turned out to be much more important than the ones they were contemplating at the beginning, task complexity increased.

Extraneous cognitive load influences an individual's ability to interact with information objects. However, no evidence was found that suggests goal-directed or exploratory behaviour can be attributed to the extraneous cognitive load imposed on an individual. The ability to be 'cognitively flexible' does not seem to significantly affect the amount of cognitive load. Without further investigation, it is difficult to suggest an explanation.

Participants who experienced high levels of extraneous cognitive load submitted a fewer number of relevance feedback passages and found it difficult to pre-determine task requirements. This could be observed when a participant identified a passage as initially not being useful but then used it later within the answer construction process.

Expertise does not seem to significantly affect the amount of cognitive load. A possible explanation is that adopting a 'schema-driven field-independent approach' does not necessarily mean that a holistic approach to the assimilation of information can always be undertaken. Individuals were not consistent in applying the same behavioural style when interacting with the information objects they encountered. It is theorised that new search heuristics are generated when an individual is unable to assimilate information during the search process.

13.2.2 KEY FINDINGS

Knowledge obtained from previous search interactions is responsible for intrinsic cognitive load, and consequently, could be partially responsible for the way individuals assimilate information. Individuals who were able to manage either 'conceptual/cognitive state' uncertainty or 'technical/affective' uncertainty in response to situational demands experienced less intrinsic cognitive load. Being able to manage intrinsic cognitive load is not associated with possessing domain expertise, applying goal-directed or exploratory behaviour, being 'cognitively flexible', or experiencing lower levels of task complexity. Interestingly, a measure of 'overall uncertainty' which incorporated both 'conceptual/cognitive state' uncertainty and 'technical/affective' uncertainty was a poorer predictor of cognitive load than when both types of uncertainty were considered independently. Depending on situational demands, a specific type of uncertainty often needs to be resolved rather than requiring both types to be resolved simultaneously. Resolving the required type of uncertainty enables individuals to assimilate information in a way that is compatible with their existing knowledge schemata. Intrinsic cognitive load is associated with the suitability of information encountered.

Task complexity is responsible for extraneous cognitive load, and consequently, could be partially responsible for the way individuals interact with information objects. Individuals who were able to spontaneously restructure knowledge in response to situational demands experienced less extraneous cognitive load. Selecting suitable heuristics that can, in some way anticipate holistic task needs, reduces task complexity and extraneous cognitive load. Being able to manage extraneous cognitive load is not associated with uncertainty, domain expertise, applying goal-directed or exploratory behaviour. The classification of problem stages encountered during information seeking enables the progress of individuals to be monitored. Exploring more subtle changes in a user's problem state, associated with task complexity, reveals that an individual's ability to pre-determine task requirements is significant. Findings showed that individuals who were able to anticipate future changes by being 'cognitively flexible' selected better heuristics. Extraneous cognitive load is associated with the suitability of the problem-solving/learning approach applied.

The model of cognitive load developed within this thesis enables the utility of feedback to be predicted during the information seeking process. Knowledge obtained from previous interactions can cause problems when assimilating (integrating) several sources of information. High levels of task complexity can cause inappropriate heuristics to be selected (problem-solving methods). Feedback has been shown to be more accurate when an individual is able to manage the cognitive load imposed by identifying 'usefulness clues'. These clues trigger new heuristics that may be more appropriate for addressing an individual's information need. Individuals undertake different information seeking (or browsing) behaviour even though they can be categorised by psychometric testing as possessing a preferred cognitive style. During the answer construction process, participants who made more modifications were likely to be engaged in a more active learning/problem-solving process. This indicated that the cognitive load imposed on them was not high enough to prevent the modification of heuristics. It is theorised that new search heuristics are generated when an individual is unable to assimilate information during the search process. These heuristics may be used to enable understanding by isolating interacting elements that cannot be understood holistically.

13.3 FURTHER RESEARCH AND RECOMMENDATIONS

The mechanism underpinning the main experimental procedure attempted to measure changes in cognitive load exclusively associated with the process of relevance feedback. An unfamiliarity associated with the process of information seeking may induce cognitive load. A within-subject experimental design was adopted in an attempt to enable cognitive load to be measured on individualistic comparative basis. However, selecting a sample that is likely to possess generally comparable cognitive capabilities, associated with IR, increases the validity of the experimental design. Therefore, it was decided that participants should be experienced users of IR systems.

Individuals selected for experimentation were involved in an information intensive work environment. Participants were either involved in researching discursive academic areas or were information professionals. The main concern of the thesis was to discover if cognitive load is measurable on an individual basis. Individuals in information intensive work environments may have developed strategies for coping with cognitive load that are more explicitly observable. The intention was not to explore comparative differences associated with different groups of individuals. However, this should represent a future research direction.

The Cognitive Load Theory proposed within this thesis has not investigated the affect of 'repetitive' or redundant information. Documents presented to experiment participants always contained conceptually different information. Future research could investigate this aspect by presenting documents that conceptually overlap. Redundant information is likely to affect both intrinsic and extraneous cognitive load. Intrinsic cognitive load is likely to increase as the suitability of information diminishes. Extraneous cognitive load is also likely to increase. When managing and predicting task requirements, redundant information is likely to increase task complexity. An individual's ability to select heuristics may be affected by having to cope with information elements that are conceptually identical.

13.3.1 IMPLICATIONS FOR THE AHRB

This study has been funded by the Arts and Humanities Research Board (AHRB). The AHRB are committed to the funding of research that allows the continuing development of information science. The AHRB are especially interested in research that enables the development of new IR systems that can intuitively adapt to user needs (AHRB, 2003). Although this research has not involved the development of an actual IR system, it has demonstrated how cognitive load affects IR. It is anticipated that future AHRB research projects can benefit from these findings. An awareness of an individual's cognitive architectural limitations should be considered when developing more intuitive IR systems (*see* Sections 11.3.3 and 11.3.4).

13.3.2 IMPLICATIONS FOR COGNITIVE LOAD RESEARCH

Cognitive Load Theory is concerned with the limitations of working memory capacity and the measures that can be taken to promote problem solving. Educational psychologists have used the theory to design more effective learning materials. However as this thesis has shown, the theory is also suitable for the investigation of information behaviour. Both causal and assessment factors affect cognitive load. Causal factors are the characteristics of an individual (e.g., cognitive abilities, and domain knowledge), while assessment factors are characteristics of the mental effort needed and the mental load imposed. Pollock, Chandler, and Sweller (2002) admitted that Cognitive Load Theory, as currently formulated, does not address the processing of highly complex information adequately, especially in language-based discursive areas such as IR. It has been found that for relevance feedback purposes, knowledge obtained from previous interactions is an important causal factor, and task complexity is an important assessment factor. The extension to Cognitive Load Theory proposed within this thesis has shown that by modelling these types of factors, it is possible to better understand the complexities associated with behaviour. The novelty of the method developed within this thesis, which allows models of behaviour derived from qualitative research to be used, provides a new approach for cognitive load researchers. Ultimately, the potential to incorporate behavioural factors may allow an insight into the way individuals manage cognitive load. Educational psychologists may be able to find ways of presenting information that supports an individual's behavioural approach to learning.

11.3.3 IMPLICATIONS FOR IR SYSTEM DEVELOPMENT

More accurate feedback judgements increase the likelihood of successful IR. This study has found that the cognitive load imposed on an individual can be assessed during information seeking. It is now theoretically possible to design IR systems in a way that identifies judgement incongruities. The goal is to develop a system capable of prompting a user to re-evaluate previous judgements that have been affected by high levels of cognitive load. However, a better understanding of information behaviour is needed before this is attainable. The experimental design developed for this thesis allows new behavioural phenomenon to be considered. Better predictors of the cognitive load imposed on an individual need to be found. A way of assessing cognitive load by observing behavioural manifestations rather than requiring a user to artificially self-report factors is required. Further research needs to find a way of incorporating a self-report mechanism within the IR process. If behavioural characteristics are identified, and the retrieval system is provided with facilities that reflect those characteristics, then the users should be able to recreate their own information seeking patterns while interacting with the system (Ellis, 1989). This type of system would allow many behavioural manifestations during information seeking to be naturally inferred.

11.3.4 IMPLICATIONS FOR INFORMATION SCIENCE

This thesis has successfully shown that an individual's cognitive architecture imposes limitations that help to explain why users are often reluctant to provide relevance feedback. The proposed Cognitive Load Theory is considerably different from Sweller's (1988) original manifestation. However, because the ideology of the original theory has been preserved, the proposed theory should be considered as an extension rather than a revision. User interaction is dependent on feedback mechanisms. The nature of these mechanisms can now be understood from a new perspective. Changes in user behaviour can be explained by the development and application of new heuristics. A lack of theory integration has meant that behavioural studies in information science have often explored different aspects of the same phenomenon. The flexibility of using Cognitive Load Theory may allow new causal and assessment factors to be identified for different types of information seeking, which can be seen as contributing to the overall mental effort associated with information processing.

AHRB, 2003. *Information science award listings.* Arts and Humanities Research Board: http://www.ahrb.ac.uk/awards/search_results.asp?panel=6

Aiken, L.S., & S.G. West, 1991. *Multiple regression: Testing and interpreting interactions.* Thousand Oaks, California: Sage Publications Inc.

Anderson, J., 1983. *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J., C. Boyle, F. Farrell & B. Reiser, 1987. Cognitive principles in the design of computer tutors. *In*: P. Morris, ed. *Modelling cognition.* NY: John Wiley.

Amabile, T. M., 1983. The Social Psychology of Creativity: A Componential Conceptualization. *Journal of Personality and Social Psychology*, **50**, 357-376.

Ayres, P., 1993. Why goal free problems can facilitate learning. *Contemporary Educational Psychology*, **18**, 376-381.

Baddeley, A.D., 1992. Working memory. *Science*, **255**, 556-559.

Baeza-Yates, R. & B. Ribeiro-Neto, 1999. *Modern Information Retrieval.* Harlow: Addison Wesley Longman Limited.

Barry, C.L., 1993. *The identification of user criteria and document characteristics: Beyond the topical approach to information retrieval.* PhD thesis, Syracuse University.

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

Bartlett, F.C., 1932. *Remembering: An experimental social study.* Cambridge: Cambridge University Press.

Bartlett, F.C., 1958. *Thinking.* New York: Basic Books.

Bates, M.J., 1986. Subject access to online catalogs: A design model. *Journal of the American Society for Information Science*, **37**, 357-376.

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

Bates, M.J., 1990. Where should the person stop and the information search interface start? *Information Processing & Management*, **26**(5), 575-591.

Bates, M.J., 1996. Document familiarity, relevance and Bradford's law: The Getty online searching project report No. 5. *Information Processing and Management*, **32**(6), 697-707.

Beaulieu, M., 1997. Experiments on interfaces to support query expansion. *Journal of Documentation*, **53**(1), 8-19.

Beaulieu, M., 2000. Interaction in information searching and retrieval. *Journal of Documentation*, **56**(4), 431-439.

Beaulieu, M. & S. Jones, 1998. Interactive searching and interface issues in the OKAPI best match probabilistic system. *Interacting with Computers*, **10**(3), 237-248.

Bechtel, W., & R.C. Richardson, 1993. *Discovering complexity.* Princeton, New Jersey: Princeton University Press.

Bedeau, M., 1997. Weak Emergence. *Philosophical Perspectives*, **11**, 375–99.

Belkin, N., 1977. *A concept of information for information science*. PhD thesis, University of London.

Belkin, N., 1978. Information concepts for Information Science. *Journal of Documentation,* **34**, 55-85.

Belkin, N., 1980. Anomalous states of knowledge as a basis for information retrieval. *The Canadian journal of information science*, **5**, 133-143.

Belkin, N., 2000. *Relevance Feedback versus Local Context Analysis as term suggestion devices: Rutgers' TREC-8 Interactive Track experience*. Rutgers University, http://trec.nist.gov/pubs.html

Belkin, N. & B. Kwasnik, 1986. Using structural representations of anomalous states of knowledge for choosing document retrieval strategies. *In*: F. Rabitti, ed. *Proceedings of the associations for computing machinery special interest group on research and development in information retrieval (Pisa, Italy).* New York: ACM press.

Belkin, N., R.N. Oddy & H.M. Brooks, 1997. ASK for information retrieval: Part I. Background and theory. *In*: K. Sparck Jones & P. Willet, eds. *Readings in Information Retrieval.* San Francisco, California: Morgan Kaufmann Publishers, 1997, pp. 299-304.

Bennett, J.L., 1972. Interactive bibliographic search as a challenge to interface design. In: D.E. Walker, ed. *Interactive bibliographic search: The user/computer interface*. Montvale, NJ: AFIPS Press.

Berry, M.W. & M. Browne, 1999. *Understanding search engines: Mathematical modelling and text retrieval.* Philadelphia: Society for Industrial and Applied Mathematics.

Boisot, M., 1998. *Knowledge Assets.* Oxford: Oxford University Press.

Boguraev, B.K., 1985. User modelling in cooperative natural language front ends. *In*: G. Gilbert & C. Heath, eds. *Social action and artificial intelligence.* London: Gower Press, pp. 124-143.

Borlund, P. & P. Ingwersen, 1997. The development of a method for the evaluation of Interactive Information Retrieval Systems. *Journal of Documentation,* **53**(3), 225-250.

Bransford, J. D. & B.S. Stein, 1984. *The Ideal Problem Solver: A Guide for Improving Thinking, Learning, and Creativity.* New York: W.H. Freeman and Company.

Brookes, B.C., 1977. The cognitive viewpoint: its development and scope. In: *CC 77: International Workshop on the Cognitive Viewpoint.* Ghent: University.

Brookes, B.C., 1980. The foundations of information science, Part 1: Philosophical Aspects. *Journal of Information Science*, **2**, 125-133.

Bruner, J., 1973. *Beyond the information given: Studies in the psychology of knowing.* New York: Norton.

Bruner, J., 1986. *Actual minds, possible worlds.* Cambridge, MA: Harvard University.

Bruner, J., 1990. *Acts of meaning.* Cambridge, MA: Harvard University.

Brünken, R., J. Plass & D. Leutner, 2002. Direct measurement of cognitive load in multimedia learning. Paper presented at: *The International Workshop on Cognitive Load Theory.* Germany: Erfurt University.

Buchdahl, G., 1969. *Metaphysics and the philosophy of science.* Oxford: Basil Blackwell, pp. 484-532.

Byström, K., & K. Järvelin, 1995. Task complexity affects information seeking and use. *Information Processing & Management,* **31**(2), 819-837.

Carter, R.F., 1974. A journalistic cybernetic. Paper presented at: *The Communication and Control of Social Processes Conference.* University of Pennsylvania.

Cartwright, N., K. Fleck, J. Cat & T. Uebel, 1994. *On Neurath's Boat.* Cambridge: Cambridge University Press.

Chang, C. & C. Hsu, 1999. Enabling Concept-Based Relevance Feedback for Information Retrieval on the WWW. *IEEE Transactions on Knowledge and Data Engineering,* **11**(4), 595-609.

Chang, S. & R.E. Rice, 1993. Browsing a Multidimensional Framework. *Annual Review of Information Science & Technology,* **28,** 619-627.

Chase, W.G. & A. H. Simon, 1973. The mind's eye in chess. *In*: W.G. Chase, ed. *Visual information processing.* New York: Academic, pp. 215-282.

Chi, M. R. Glaser & M. Farr, 1988. *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Clark-Carter, D., 1998. *Doing quantitative psychological research: From design to report.* Hove, Sussex: Psychology Press, Taylor & Francis.

Colby, K.M., 1973. Simulations of belief systems. *In*: R.C Shank & K.M Colby, eds. *Computer models of thought and language.* San Francisco: Freeman, pp. 251-286.

Cooper, G., 1990. Cognitive load theory as an aid for instructional design. *Australian Journal of Educational Technology,* **6**(2), 108-113.

Cooper, G., 1998. *Research into Cognitive Load Theory and Instructional Design at UNSW.* Sydney, Australia: University of New South Wales, http://www.arts.unsw.edu.au/eductaion/CLT_NET_Aug_97.html

Cooper, G. & J. Sweller, 1987. Effects of schema acquisition and rule automation on mathematical problem solving transfer. *Journal of Educational Psychology*, **79**, 347-362.

Cove, J.F. & B.C. Walsh, 1988. Online text retrieval via browsing. *Information Processing and Management*, **24**(1), 31-37.

Daniels, P.J., 1986. Cognitive models in information retrieval: An evaluative review. *Journal of Documentation*, **45**(4), 272-304.

Daniels, P.J., H.S. Brooks, & N.J Belkin, 1997. Using problem structures for driving human-computer dialogues. *In*: K. Sparck Jones & P. Willet, eds. *Readings in Information Retrieval.* San Francisco, California: Morgan Kaufmann Publishers, pp. 135-142.

De May, M., 1977. The cognitive viewpoint: its development and scope. *In: CC 77: International Workshop on the Cognitive Viewpoint.* Ghent University.

Dent, E.B., 1999. Complexity Science: A worldview shift. *Emergence*, **1**(4), 5-19.

Dervin, B., 1983. *An overview of sense-making research: Concepts, methods, and results to date*. Seattle: School of Communications, University of Washington.

Dervin, B. & M. Nilan, 1986. Information needs and uses. *Annual Review of Information Science and Technology*, **21**, 3-33.

Dewey, J., 1933. *How we think.* Lexington, MA: Heath.

Doyle, L.B, 1963. Is relevance an adequate criterion in retrieval system evaluation. *In*: *Proceedings of the American Documentation Institute 26th Annual Meeting.* Washington DC: American Documentation Institute, pp.199-200.

Efthimiadis, E.N., 1992. *Interactive query expansion and relevance feedback for document retrieval systems*. PhD thesis, London: City University.

Efthimiadis, E.N., 1996. Query Expansion. *Annual Review of Information Science and Technology*, **31**, 121-187.

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

Ellis, D., 1984. Theory and explanation in information retrieval. *Journal of Information Science*, **8**, 25-38.

Ellis, D., 1989. A behavioural 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.

Ellis, D., N. Ford, & F. Wood, 1993. Hypertext and learning styles. *The Electronic Library*, **11**, 13-18.

Entwistle, N.J., 1981. *Styles of learning and teaching.* London: Wiley.

Excite Inc., 1996. *Information retrieval technology and Intelligent Concept Extraction searching.* http://www.excite.com/ice/tech.html

Fauconnier, G., 1997. *Mappings in thought and language.* Cambridge: Cambridge University Press.

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

Festinger, L., 1957. *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Feyerabend, P., 1993. *Against Method.* 3rd Edition. London: Verso.

Finke, R., S. Smith, & T. Ward, 1996. *Creative Cognition: Theory, Research, and Applications.* Cambridge, MA: MIT Press.

Flavell, J.H., 1963. *The developmental psychology of Jean Piaget.* New York: Van Nostrand Reinhold.

Ford, N., 1999. The growth of understanding in information science: towards a developmental model. *Journal of the American Society for Information Science*, **50**(12), 1141-1152.

Ford, N., 2000. Cognitive styles and virtual environments. *Journal of the American Society for Information Science*, **51**(6), 543-557.

Foss, C.L., 1989. Tools for reading and browsing hypertext. *Information Processing and Management*, **25**(4), 407-418.

Fransson, A., 1984. Cramming or understanding? Effects of intrinsic and extrinsic motivation on approach to learning and test performance. *In*: J. Alderson & A. Urquhart, eds. *Reading in a foreign language*. White Plains, NJ: Longman.

Gagne, R., 1970. *The conditions of learning*. 2nd Edition. New York: Holt, Rinehart & Winston.

Gerjets, P. & K. Scheiter, 2002. Goal configurations and processing strategies as mediators between instructional design and cognitive load. Paper presented at: *The International Workshop on Cognitive Load Theory*. Germany: Erfurt University.

Gentner D. & A. Stevens, 1983. *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ghiselin, B., 1952. *The creative process*. New York: New American Library.

Goldstein, J., 1996. Causality and Emergence in Chaos and Complexity Theories. *In:* W. Sulis & A. Combs, eds. *Nonlinear Dynamics and Human Behaviour*. Singapore: World Scientific Publishing, pp. 161–90.

Goldstein, J., 1999. Emergence as a Construct: History and Issues. *Emergence*, **1**(1), 49-72.

Greeno, J.G., 1974. Hobbits and Orcs: Acquisition of a sequential concept. *Cognitive Psychology*, **6**, 270-292.

Guilford, J.P., 1956. The Structure of Intellect. *Psychological Bulletin*, **53**, 267-293.

Guilford, J.P., 1968. *Intelligence, creativity, and their educational implications*. San Diego, CA.: R.R. Knapp.

Hacking, I., 1975. *The emergence of probability: A philosophical study of early ideas about probability, induction, and statistical inference*. London: Cambridge University Press.

Harman, D., 1992. Relevance feedback and other query modification techniques. *In:* W.B. Frakes & R. Baeza-Yates, eds. *Information Retrieval Data Structures and Algorithms*. Englewood Cliffs, New Jersey: Prentice Hall, pp. 241-263.

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.

Harter, S.P., & C.A. Hert, 1997. Evaluation of information retrieval systems: approaches, issues and methods. *Annual Review of Information Science and Technology (ARIST)*, **32**, 3-94.

Hayes, J.R & H. Simon, 1976. Understanding complex task instruction. *In*: D. Klahr, ed. *Cognition and instruction*. Hillsdale, NJ: Erlbaum, pp. 269-286.

Heinström, J., 2002. Fast surfers, broad scanners and deep divers. Personality and information-seeking behaviour. PhD thesis, Turku (Åbo): Åbo Akademi University.

Heit, E., 1997. Knowledge, concepts, and categories. *In*: K. Lamberts & D. Shanks, eds. *Knowledge and concept learning*. Hove: Psychology Press, pp. 7-41.

Henle, P., 1942. The Status of Emergence. *Journal of Philosophy*, **39**, 486–93.

Hert, C. A., 1997. *Understanding Information Retrieval Interactions: Theoretical and Practical Implications*. Greenwich, Connecticut: Ablex Publishing Corporation.

Hidi, S., 1990. Interest and its contribution as a mental resource for learning. *Review of Educational Research*, **60**, 549-571.

Höglund, L., 2003. Review of: Heinström, Jannica. Fast surfers, broad scanners and deep divers. Personality and information-seeking behaviour. Turku (Åbo): Åbo Akademi University Press, 2002. *Information Research*, **8**(2).

Holland, J., 1998. *Emergence: from Chaos to Order*. Reading, MA: Addison-Wesley.

Hu, P.J., P.C. Ma & P.Y. Chau, 1999. Evaluation of user interface designs for information retrieval systems: a computer-based experiment. *Decision Support Systems*, **27**(1-2), 125-143.

Huber, G.L., J.H.W Roth, 1999. *Finden oder suchen? Lehren und Lernen in Zeiten der Ungewißheit*. [Finding or Searching? Teaching and learning in uncertain times] Schangau: Ingeborg Huber.

Infoseek Corporation, 1998. *Infoseek introduces "E.S.P." to dramatically improve general search results*. http://www.ir-infoseek.com/1998_releases/esp.phtml

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

Ingwersen, P., 1992. *Information Retrieval Interaction.* London: Taylor Graham Publishing.

Ingwersen, P., 1996. Cognitive perspectives of Information Retrieval Interaction: Elements of a cognitive IR theory. *Journal of Documentation*, **52**(1), 3-50.

Ingwersen, P., 1999. Cognitive information retrieval. *Annual Review of Information Science and Technology*, **34**, 3-52.

Iselin, E., 1989. The impact of information diversity on information overload effects in unstructured managerial decision making. *Journal of Information Science*, **15**(1), 163-173.

Janes, J.W., 1989. The application of search theory to information science. *Proceedings of the 52nd annual meeting of the American Society for Information Science (ASIS)*, **26**, 9-12.

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

Johnson-Laird, P.N., 1983. *Mental models.* Cambridge: Harvard University.

Jonassen, D., D. Ambruso & J. Olesen, 1992. Designing hypertext on transfusion medicine using cognitive load theory. *Journal of Educational Multimedia and Hypermedia*, **1**(3), 309-322.

Kauffman, S., 1995. *At Home in the Universe: the Search for the Laws of Self-Organization and Complexity.* New York: Oxford University Press.

Kelly, G.A., 1963. A theory of personality: The psychology of personal constructs. New York: Norton.

Kirschner, P.A., 2002. Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction*, **12**, 1-10.

Kochen, M., 1983. Library science and information science. *In:* F. Machlup & U. Mansfield, eds. *The study of information.* New York: Wiley & Sons, pp. 371-377.

Kolb, D.A., 1984. *Experiential learning*. Englewood Cliffs, NJ: Prentice-Hall.

Kotovsky, K., J.R. Hayes, H. Simon, 1985. Why are some problems hard? Evidence from tower of Hanoi. *Cognitive Psychology*, **17**, 248-294.

Kuhlthau, C., 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, **42**(5), 361-371.

Kuhlthau, C., 1993a. A principle of uncertainty for information seeking. *Journal of Documentation*, **49**(4), 339-355

Kuhlthau, C., 1993b. *Seeking meaning: A process approach to library and information services*. Norwood, NJ: Ablex.

Kuhlthau, C., A. Spink & C. Cool, 1992. Exploration into stages in the information search process in online information retrieval. *Proceedings of the annual meeting of the American Society for Information Science (ASIS)*, **29**, 67-71.

Kuhn, T.S., 1996. *The structure of scientific revolutions*. 3rd ed. Chicago: University of Chicago Press.

Laird, J.E., A. Newell & P.S. Rosenbloom, 1987. Soar: An architecture for general intelligence. *Artificial Intelligence*, **33**, 1-64.

Larkin, H., J. McDermott, D. Simon & H. Simon, 1980. Models of competence in solving physics problems. *Cognitive Science*, **11**, 65-99.

Lefebvre, E., & H. Letiche, 1999. Managing Complexity from Chaos: Uncertainty, Knowledge and Skills. *Emergence,* **1**(3), 7-15.

Lewes, G.H., 1875. *Problems of Life and Mind*. Vol. 2. London: Kegan, Paul, Trench, Turbner.

Marchionini, G., 1987. An invitation to browse: Designing full text systems for novice users. *The Canadian journal of information science*, **12**(3-4), 69-79.

MacCall, S.L, 1998. Relevance reliability in cyberspace: towards measurement theory for Internet information retrieval. Information access in the global information economy. *Proceedings of the 61st annual meeting of the American Society for Information Science (ASIS)*, **35**, 13-22.

McLeod, J. & Cropley, 1989. *Fostering academic excellence*. Pergamon.

McClelland, J.L. & D.E. Rumelhart, 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 2, Psychological and biological Models).* Cambridge, MA: MIT Press.

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, G.A., E. Galanter & K.H. Pribram, 1960. *Plans and the structure of behaviour*. New York: Holt, Rinehard & Winston.

Moens, M., 1999. Personalised Information Objects. Paper presented at: *IEE Infomatics seminar on searching for information: Artificial intelligence and information retrieval approaches*. Scotland: Glasgow.

Moran, T.P., 1981. Command language grammar: A representation scheme for the user interface of interactive systems. *International Journal of Man-Machine Studies*, **15**(1), 3-50.

Newell, A., 1973. Production systems: Models of control structures. *In*: W.G. Chase, ed. *Visual information processing*. New York: Academic, pp. 463-526.

Oddy, R.N., 1977. Information retrieval through man-machine dialogue. *Journal of Documentation*, **33**, 1-14.

Oddy, R.N. & B. Balakrishnan, 1991. PTHOMAS: An adaptive information retrieval system on the connection machine. *Information Processing and Management*, **27**(4), 317-335.

Overmeer, M., 1999. A search interface for my questions. *Computer Networks*, **31**(21), 2263-2270.

Paas, F., 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, **84**, 429-434.

Paas, F. & J. Van Merrienboer, 1994. Variability of worked examples and transfer of geometric problem-solving skills: A cognitive load approach. *Journal of Educational Psychology*, **86**, 122-133.

Palmquist, R.A. & K. Kim, 2000. Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science*, **51**(6), 558-566.

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

Pask, G., 1976a. Styles and strategies of learning. *British Journal of Educational Psychology*, **46**, 128-148.

Pask, G., 1976b. *Conversation theory: Applications in education and epistemology.* Amsterdam: Elsevier.

Peat, H.J. & P. Willett, 1991. The limitation of term co-occurrence data for query expansion in document retrieval systems. *Journal of the American Society for Information Science*, **42**(5), 378-383.

Pepper, S., 1926. Emergence. *Journal of Philosophy*, **23**, 241–245.

Perkins, D.N., 1981. *The Mind's Best Work.* Cambridge, MA: Harvard University Press.

Pollock, E., P. Chandler & J. Sweller, 2002. Assimilating complex information. *Learning and Instruction*, **12**, 61-89.

Popper, K., 1959. *The logic of scientific discovery.* London: Hutchinson.

Proper, H.A. & P.D. Bruza, 1999. What is information discovery about? *Journal of the American Society for Information Science*, **50**(9), 737-750.

Rich, E.A., 1983. Users are individuals: Individualising user models. *International Journal of Man-Machine Studies*, **18**, 199-214.

Richardson, R.C., 1986. Models and scientific explanations. *Philosophica*, **37**, 59-72.

Rieger, C., 1976. An organisational knowledge for problem solving and language comprehension. *Artificial Intelligence*, **7**, 89-127.

Robertson, S.E., 1990. On term selection for query expansion. *Journal of Documentation*, **46**(4), 359-364.

Robertson, S.E. & M. Hancock-Beaulieu, 1992. On the evaluation of IR systems. *Information Processing and Management*, **28**(4), 219-236.

Rocchio, J., 1971. Relevance Feedback in Information Retrieval. *In*: G. Salton, ed. *The SMART Retrieval Storage and Retrieval System.* Englewood Cliffs, New Jersey: Prentice Hall, pp. 313-323.

Salton, G. & Buckley, C., 1997. Improving retrieval performance by Relevance Feedback. *In*: K. Sparck Jones & P. Willet, eds. *Readings in Information Retrieval.* San Francisco, California: Morgan Kaufmann Publishers, pp. 355-364.

Saracevic, T., 1996. Modelling interaction in information retrieval: A review and proposal. *Annual Review of Information Science and Technology (ARIST)*, 3-9.

Saracevic, T., 1997. Relevance: A review of and Framework for the thinking on the notion in Information Science. *In*: K. Sparck Jones & P. Willet, eds. *Readings in Information Retrieval.* San Francisco, California: Morgan Kaufmann Publishers, pp. 143-165.

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

Saracevic, T., P. Kantor, A. Chamis. & D. Trivison, 1997. A study of information seeking and retrieving: Background and methodology. *In*: K. Sparck Jones & P. Willet, eds. *Readings in Information Retrieval.* San Francisco, California: Morgan Kaufmann Publishers, pp. 175-190.

Savage-Knepshield, P.A. & N.J. Belkin, 1999. Interaction in information retrieval: Trends over time. *Journal of the American Society for Information Science*, **50**(12), 1067-1082.

Schamber, L., 1991. Users' criteria for evaluation in a multimedia environment. *Proceedings of the annual meeting of the American Society for Information Science (ASIS)*, **28**, 27-31.

Schamber, L., 1994. Relevance and information behaviour. *Annual Review of Information Science and Technology*, **29**, 3-48.

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

Schneider, W. & R.M. Shiffrin, 1977. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, **84**, 1-66.

Schutz, A., 1967. *The phenomenology of the social world.* Evanston, IL: Northwestern University Press.

Searle, J., 1994. *The Rediscovery of Mind (Representation and Mind)*, New York: Bradford Books.

Shannon, C. & W. Weaver, 1949. *The mathematical theory of communication.* Urbana, IL: University of Illinois Press.

Shouksmith, G., 1970. *Intelligence, Creativity and Cognitive Style.* New York: Wiley-Interscience.

Simon, H., 1969. *The sciences of the artificial.* 2nd Edition. Cambridge: M.I.T. Press.

Simon, H., 1973. The structure of ill-structured problems. *Artificial Intelligence*, **4**, 181-201.

Simon, H., 1981. Information-processing models of cognition. *Journal of the American Society for Information Science*, **32**, 364-377.

Simon, H. & S.K. Reed, 1976. Modelling strategy shifts in problem solving task. *Cognitive Psychology*, **8**, 86-97.

Sorrentino, R. M., D.R. Bobocel, M.Z Gitta, J. Olson & E.C. Hewitt, 1988. Uncertainty orientation and persuasion: Individual differences in the effects of personal relevance on social judgements. *Journal of Personality and Social Psychology*, **55**, 357-371.

Sparck Jones, K., S. Walker & S.E. Robertson, 1998. *A Probabilistic Model of Information Retrieval: Development and Status (TR 446).* University of Cambridge: Computer Laboratory, http://www.cl.cam.ac.uk

Spink, A., 1996. Interactive information seeking and retrieving: A third feedback framework. *Proceedings of the annual meeting of the American Society for Information Science (ASIS)*, **33**, 10-15.

Spink, A., 1997. Information Science: A third feedback framework. *Journal of the American Society for Information Science*, **48**(8), 728-740.

Spink, A., H. Greisdorf, 2001. Regions and Levels: Mapping and Measuring Users' Relevance Judgments. *Journal of the American Society for Information Science*, **52**(13), 161-173.

Spink, A., H. Greisdorf & J. Bateman, 1998. From highly relevant to not relevant: Examining different regions of relevance. *Information Processing and Management*, **34**(5), 599-621.

Spink, A. & R.M Losee, 1996. Feedback in information retrieval. *Annual Review of Information Science and Technology*, **31**, 33-78.

Spink, A. & T. Saracevic, 1997. Interaction in information retrieval: Selection and effectiveness of search terms. *Journal of the American Society for Information Science*, **48**(8), 741-761

Spink, A. & T. Saracevic, 1998. Human-computer interaction in information retrieval: nature and manifestations of feedback. *Interacting with Computers*, **10**(3), 249-267.

Spiro, R.J. & J. Jehng, 1990. Cognitive flexibility and hypertext: Theory and technology for the non-linear and multidimensional traversal of complex subject matter. *In*: D. Nix & R. Spiro, eds. *Cognition, education and multimedia.* Hillsdale, NJ: Erlbaum.

Stark, R., H. Mandl, H. Gruber & A. Renkl, 2002. Conditions and effects of example elaboration. *Learning and Instruction*, **12**, 39-60.

Sullivan, D., 2000. *Search Engine Watch™, Mecklermedia.* http://www.searchenginewatch.com/

Sweller, J., 1993. Some cognitive-processes and their consequences for the organisation and presentation of information. *Australian Journal of Psychology*, **45**(1), 1-8.

Sweller, J., 1994. Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, **4**, 295-312.

Sweller, J., 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, **12**, 257-285.

Sweller, J., 1999. *Instructional design in technical areas.* Melbourne: ACER.

Sweller, J. & P. Chandler, 1991. Evidence for cognitive load theory. *Cognition and Instruction*, **8**(4), 351-362.

Sweller, J., P. Chandler, P. Tierner & M. Cooper, 1990. Cognitive load in the structuring of technical material. *Journal of Experimental Psychology: General,* **119,** 176-192.

Sweller, J., J. Van Merrienboer & F. Paas, 1998. Cognitive Architecture and Instructional Design. *Educational Psychology Review,* **10**(3), 251-296.

Swets, J.A., 1963. Information retrieval systems. *Science,* **141**, 245-250.

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

Taylor, R.S., 1986. *Value added processes in information systems.* Norwood, NJ: Ablex.

Thomas, N.P., 1993. Information seeking and the nature of relevance. *Proceedings of the annual meeting of the American Society for Information Science (ASIS),* **30,** 126-130.

Thompson, R.H. & B.W. Croft, 1989. Support for browsing in an intelligent text retrieval system. *International journal of man-machine studies,* **30**, 639-668.

Tuovinen, J.E. & J. Sweller, 1999. A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology,* **91**(2), 334-341.

Van Merriënboer, J., P.A Kirschner & L. Kester, 2002. Just-in-time information presentation and the acquisition of complex cognitive skills. Paper presented at: *The International Workshop on Cognitive Load Theory.* Germany: Erfurt University.

Van Merrienboer, J. & H. Krammer, 1987. Instructional strategies and tactics for the design of introductory computer programming courses in high school. *Instructional Science,* **16**, 251-285.

Van Rijsbergen, C.J., 1979. *Information Retrieval.* 2nd ed. London: Butterworth.

Vakkari, P., 1999. Task complexity, problem structure, and information actions. Integrating studies on information seeking and retrieval. *Information Processing & Management,* **35**, 819-837.

Walker, S. & R. De Vere, 1990. *Improving subject retrieval in Online Catalogues: Relevance feedback and query expansion.* London: British Library Research Paper 72.

Watkins, M.J. & E. Tulving, 1975. Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, **104**, 5-29.

Wersig, G., 1971. *Information – Kommunikation – Dokumentation.* Pullach bei München: Verlag Dokumentation.

Whittemore, B & M. Yovits, 1974. A generalized conceptual development for the analysis and flow of information. *Journal of the American Society for Information Science*, **24**, 221-231.

Wilson, T., 1981. On user studies and information needs. *Journal of Documentation*, **37**, 3-15.

Wilson, T., 1999a. Exploring models of information behaviour: The uncertainty project. *Information Processing and Management*, **35**, 839-849.

Wilson, T., 1999b. Models of information behaviour research. *Journal of Documentation*, **55**, 249-270.

Wilson, T., N. Ford, D. Ellis, A.E. Foster & A. Spink, 2000. Uncertainty and its correlates. Paper presented at: *Information Seeking in Context.* Sweden, Göthenburg. See also: *The New Review of Information Behaviour Research*, **1**, 69-84.

Wimsatt, W., 1976. Reductionism, Levels of Organization and the Mind-Body Problem. *In*: G. Globus, G. Maxwell & I. Sabodnik, eds. *Consciousness and the Brain*, New York: Plenum Press.

Winograd, T. & F. Flores, 1986. *Understanding computers and cognition: A new foundation for design.* Norwood, NJ: Ablex.

Wisniewski, E. J. & D. Gentner, 1991. On the combinatorial semantics of noun pairs: minor and major adjustments to meaning. *In*: G.B. Simpson, ed. *Understanding Word and Sentence*, North Holland: Elsevier, pp. 241-284.

Witkin, H.A. & D.R. Goodenough, 1981. *Cognitive styles, essence and origin: Field dependence and field independence.* New York: International Universities Press, Inc.

Witkin, H.A., C.A. Moore, D.R. Goodenough & P.W. Cox, 1977. Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, **47**, 1-64.

Yoon, K., & M.S. Nilan, 1999. Towards a reconceptualization of information seeking research: focus on the exchange of meaning. *Information Processing and Management*, **35**, 871-890.

Yovits, M. & C. Foulk, 1985. Experiments and analysis of information use and value in a decision making context. *Journal of the American Society for Information Science*, **36,** 63-81.

TRADEMARKS

AltaVista™, Overture Services, Inc. http://www.altavista.com/

DT Search™ Version 5.1, DT Search Corp. http://www.dtsearch.com

Encyclopaedia Britannica™.
http://www.britanica.com

Excite™, Excite Inc. http://www.excite.com/

Google™, Google. http://www.google.com/

Infoseek™, Infoseek Corporation. http://www.infoseek.com/ (defunct)

Kenjin™, Autonomy. http://www.autonomy.com/

Lycos™, Lycos Inc, Carnegie Mellon University. http://www.lycos.com/

Microsoft Visual Basic 6.0™, Microsoft Corporation 2003.

SPSS Version 10.0™, SPSS Inc. 2003.

WebCrawler™, InfoSpace Inc. http://www.webcrawler.com/

APPENDIX A

E-mail:

Subject: Re: Feedback Research
From: j.back@lboro.ac.uk
To: jb-volunteer-list@lboro.ac.uk

Dear Participant,
Thank you for agreeing to participate.
There are two straightforward stages that you need to complete.
Instructions are included within this email (save this email for future reference).
If you have any questions please contact me: j.back@lboro.ac.uk

Stage 1: Submit your query to me
Use your e-mail software to reply to this email (reply to sender).
Please Answer the following 2 questions:

Q1. Please write a sentence that explains the topic area you have chosen for your final year project.

My final year project investigates...

Q2. Please compile a list of four keywords that best explain this topic area.

1.
2.
3.
4.

Now send your reply. You will get a response within 24 Hours.
I will send you a software attachment via email.
On receipt please save the attachment.

Stage 2: Feedback
Once you have received & saved the attachment please run the software.
You will be presented with five Web pages for evaluation.
Please follow on-screen instructions.
You will be required to provide three types of feedback for each Web page:

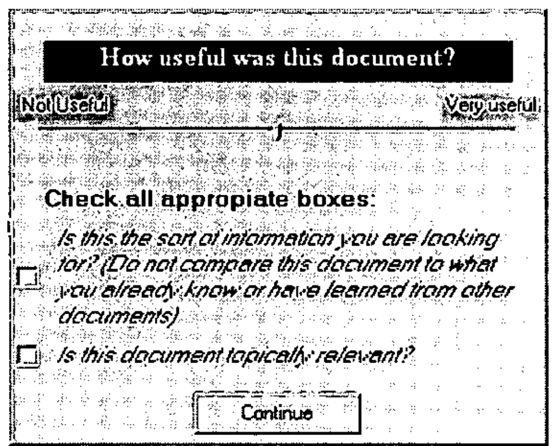
1. "Is this the sort of thing you are looking for? - Do not compare this page with what you already know or have learned from other pages."
2. "How useful is this page? - Compare this page with what you already know or have learned from other pages."
3. Using the highlighting tool please highlight passages that you find to be useful.

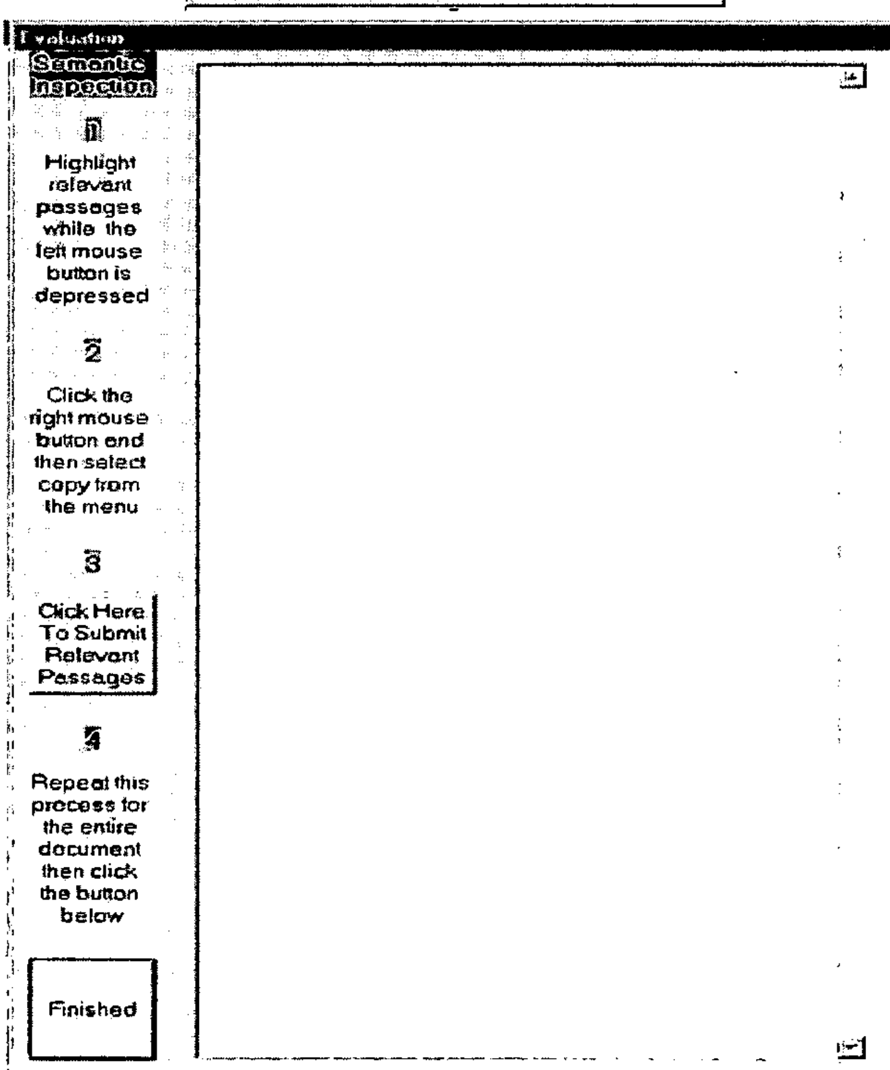
The feedback process is straightforward and is unlikely to take more than fifteen minutes. On completion the software will automatically send your responses.

Thank you very much for your time and effort.

Software:

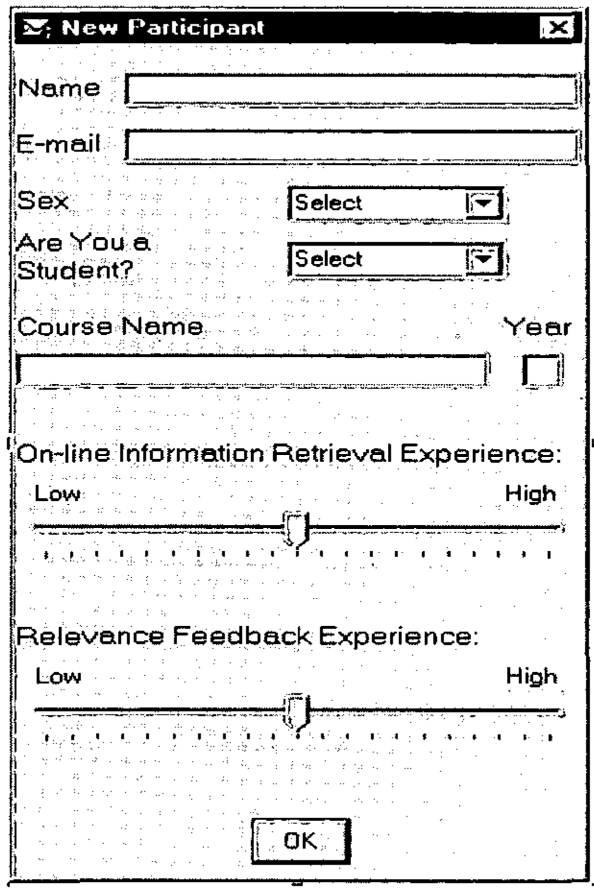
The screenshots below show how the three types of feedback were submitted by participants. $RF_{a0.75}$ and $RF_{a0.25}$ simulated the check-box approach. A more spontaneous method of obtaining judgments was used for RF_b. While users browsed the five documents returned, they were asked to mark what they considered to be the most useful passages using a highlighting tool.



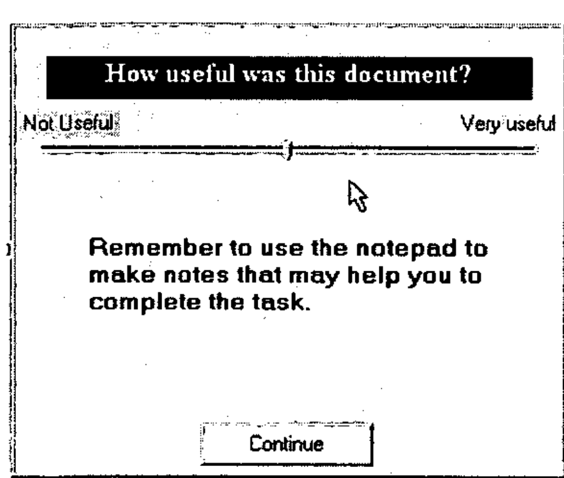


APPENDIX B

Questionnaire that gathered background data:



Six documents (HTML pages with no hyperlinks) were presented to users in a random order. Each participant was asked to evaluate the usefulness of the document. This evaluation was performed using a **magnitude scale:**



APPENDIX C

Questionnaire:

Part A Please respond on a scale from 1-7

1 = Always affects my approach*
2 = Frequently affects my approach*
3 = Generally **does** affect my approach*
4 = Sometimes affects my approach
5 = Generally **does not** affect my approach
6 = Rarely affects my approach
7 = Never affects my approach

* If you answer 1, 2 or 3, please explain how your approach is affected:

What do you do differently?:

A) Do you highlight more or less passages?	More	Less	Don't know
B) Do you select your passages more carefully?	Yes	No	Don't know

1. If you are **uncertain** that highlighting a passage will provide good results does this affect your highlighting approach?
2. If you are **unfamiliar** with the subject area you are searching does this affect your highlighting approach?
3. If you have searched a subject area before and **failed to find useful information** does this affect your highlighting strategy when performing another search on the same subject area?
4. If you consider yourself to have **good knowledge** about a subject area you **regularly** search does this affect your highlighting strategy?

Part B Free Response

What factors affect your willingness to highlight a passage:

1. When do you find it most difficult to highlight a passage?
2. When do you find it easiest to highlight a passage?

APPENDIX D

During the main experimental procedure participants were allowed to undertake any two tasks from a pool of seven.

Task format:
Adapted from Borlund and Ingwersen's (1997) simulated information need design.

Indicative request:	A general request for information. E.g. "Find something about…".
Definition:	Key concept definitions including details on associations between concepts. E.g. "Concept X explains Concept Y if Concept Z is False".
Simulated situation:	Source of need, environmental situation, and the problem that is required to be addressed.

Task 1	
Indicative request:	Find information about carboxylic acid.
Definition:	Joseph Lister faced staunch criticism from the medical community and the public as he lost a number of nurses to respiratory failure by using carboxylic acid during surgery.
Simulated situation:	Medieval medicine professed that all illness was caused by evil spirits and they should be cast out by blood-letting and exorcisms. What led Joseph Lister to think otherwise? How could he have known he was right?

Task 2

Indicative request: Find information about the heliocentric viewpoint.

Definition: In 260 BC Aristarchos put forward the heliocentric theory of the Solar System, which puts the Sun at the centre of the system of the six planets then known, with the stars infinitely distant. Aristarchos maintained that the planets moved in Platonically perfect circles.

Simulated situation: Why did Copernicus criticise the Ptolemaeans? Why was Galileo's later work on tides helpful to his ideas on the solar system but wrong scientifically?

Task 3

Indicative request: Find information about the age of the Earth.

Definition: The physicist Kelvin was thought to be correct in his estimates for the age of the Earth even though there was much evidence from other disciplines against his theory.

Simulated situation: Why did so many scientists agree with him?

Task 4

Indicative request: Find information about the Phlogiston theory.

Definition: Stahl suggested that phlogiston was weightless or could even have negative weight.

Simulated situation: Priestly and Laviosier were opponents of the phlogiston theory but found it difficult to produce contrary evidence.
Why was it difficult to produce this evidence?

Task 5

Indicative request: Find information about Rationalists and Positivists.

Definition: Rationalists suggest that ideas should be developed in a way that allows them to be easily criticised. However, according to the notion of incommensurability, there is no neutral or objective standpoint, and therefore no rational and objective way in which one particular theory can be chosen over another.

Simulated situation: Was Sir Karl Popper a Rationalist?
Was Paul Feyerabend a Rationalist?

Task 6

Indicative request: Find information about Quantum theory.

Definition: Aspect and his team found that if one of a pair of photons was polarized in a certain way, its twin at the other end of the apparatus would always be polarized in the same way. It is as if the photons know what is happening to each other, even though there can be no possible communication between them.

Simulated situation: Was Albert Einstein proved to be right or wrong?

Task 7

Indicative request: Find information about Alzheimer's disease.

Definition: Various factors have been implicated in causing Alzheimer's disease including high levels of aluminium in drinking water and the presence in the brain of an abnormal protein, known as beta-amyloid.

Simulated situation: Why are scientists finding it difficult to find a cure?

Anticipated generation of cognitive load

Task 1:

Participants needed to recognise that Lister was influenced by the work of Pasteur. Learning of Pasteur's discovery of micro-organisms, Lister began to use carbolic acid as a disinfectant. This connection was not explicitly made within the documents presented and had to be inferred.

Task 2:

Copernicus' heliocentric system implied anomalies that could not be explained within the physical explanations available in contemporary Aristotelian philosophy. Participants were required to infer the link between developments in philosophical thinking and scientific observation. This connection was not explicitly made within the documents presented and had to be inferred.

Task 3:

Participants needed to recognise that Darwin's work was seen as controversial explaining why Kelvin's ideas were more popular at the time. This connection was not explicitly made within the documents presented and had to be inferred.

Task 4:

Lavoisier finally drew all the evidence together, and discovered the oxygen theory of combustion. He did not come to his conclusions immediately, but as the result of several related steps. Chemists had no clear idea of what a chemical element was, nor any understanding of the nature of gases. The relations between these steps was not explicitly made within the documents presented and had to be inferred.

Task 5:

Documentation included information on Popper's views and various philosophical viewpoints including 'rationalism'. Science, in Popper's view, starts with problems rather than with observations. This is a typical 'rationalist' viewpoint (Feyerabend was not a 'rationalist'). Popper was not explicitly labelled as a 'rationalist' within the documents. This had to be inferred.

Task 6:

Documentation included information on Einstein's theory of relativity. Einstein could not accept the idea that in the subatomic world events happen almost by chance, and particles do not have exact positions. Therefore, Aspect's findings proved Einstein to be wrong. This connection was not explicitly made within the documents presented and had to be inferred.

Task 7:

Documentation included a wide range of information on treatments. Several factors have to come together in a person before Alzheimer's disease develops. A cure will not be found until more is known about these factors. This connection was not explicitly made within the documents presented and had to be inferred.

APPENDIX E

An example of a log file created during the main experimental procedure
(Annotations in italics)

LOG FILE ID: 00189
DATE: 08/08/2002
LOG START TIME: 14:04
LOG END TIME: 14:37
PARTICIPANT ID: 0024

BACKGROUND DATA COLLECTION *See Section 10.3*

NAME: Steve Grey
SEX: M
AGE: 26
OCCUPATION: software engineer
ON-LINE RETRIEVAL EXPERIENCE: 0.85 *Magnitude Scale [Min= 0 Max= 1]*
RELEVANCE FEEDBACK EXPERIENCE: 0.32

TASK SELECTION *See Appendix D*

TASK ID: 5
INDICATIVE REQUEST: Find information about Rationalists and Positivists.
DEFINITION: Rationalists suggest that ideas should be developed in a way that allows them to be easily criticised. However, according to the notion of incommensurability, there is no neutral or objective standpoint, and therefore no rational and objective way in which one particular theory can be chosen over another.
SIMULATED SITUATION:Was Sir Karl Popper a Rationalist? Was Paul Feyerabend a Rationalist?

INTRINSIC MEASURE OF COMPARISON

INITIAL USEFULNESS JUDGEMENT *See Section 10.3.2*
DOC 1: 0.72 *Magnitude Scale [Min= 0 Max= 1]*
DOC 2: 0.65
DOC 3: 0.41
DOC 4: 0.84
DOC 5: 0.50

USEFULNESS JUDGEMENT *See Section 10.3.4*
AFTER FINAL ANSWER CONSTRUCTION
DOC 1: 0.48 *Magnitude Scale [Min= 0 Max= 1]*
DOC 2: 0.70
DOC 3: 0.88
DOC 4: 0.65
DOC 5: 0.24

The variance between an initial usefulness (iuj) judgement and a usefulness judgement after final answer construction (puj) is used as the intrinsic measure of comparison (imc) once a baseline measure is calculated (bicl) (See Section 9.1)

VARIABLES CAPTURED IN ORDER TO EXPLAIN INTRINSIC MEASURE OF COMPARISON VARIANCE *See Section 10.3.1*

PRE-TEST	*Magnitude Scale [Min= 0 Max= 1]*
CON/COG_UNC: 0.21	*Conceptual Cognitive State Uncertainty*
TEC/AFF_UNC: 0.48	*Technical Affective State Uncertainty*
DOM_KNO: 0.15	*Domain Knowledge*
TYP_KNO: 0.52	*Background or Specific Information*

AFTER DOCUMENT 1
CON/COG_UNC: 0.46
TEC/AFF_UNC: 0.25
DOM_KNO: 0.35
TYP_KNO: 0.46

AFTER DOCUMENT 2
CON/COG_UNC: 0.45
TEC/AFF_UNC: 0.65
DOM_KNO: 0.32
TYP_KNO: 0.65

AFTER DOCUMENT 3
CON/COG_UNC: 0.54
TEC/AFF_UNC: 0.85
DOM_KNO: 0.31
TYP_KNO: 0.49

AFTER DOCUMENT 4
CON/COG_UNC: 0.45
TEC/AFF_UNC: 0.90
DOM_KNO: 0.65
TYP_KNO: 0.78

AFTER DOCUMENT 5
CON/COG_UNC: 0.68
TEC/AFF_UNC: 0.71
DOM_KNO: 0.56
TYP_KNO: 0.68

AFTER DOCUMENT 6
CON/COG_UNC: 0.81
TEC/AFF_UNC: 0.95
DOM_KNO: 0.61
TYP_KNO: 0.45

TASK FULFILMENT EVALUATION: 0.84 *See Section 10.3.4*

The self reported variables captured above (see Section 9.5) are used to populate a model of intrinsic cognitive load by predicting values of the intrinsic measures of comparison (imc) (see Section 9.4 and 11.9)

EXTRANEOUS COGNITIVE LOAD EVALUATION

See Section 10.3.2 and 10.3.3

DOCUMENT 1

Feedback passages selected

P001: who rejected the attempt by certain philosophers to find a methodology applicable to all scientific research.
P002: scientific progress is claimed to be the result of a range of sociological factors working to promote politically convenient notions of how nature operates.
P003: His philosophy has been interpreted as 'anything goes',

ANSWER CONSTRUCTION:
Popper is a philosopher. Feyerabend [P001: rejected the attempt by certain philosophers to find a methodology applicable to all scientific research.] Feyerabend's [P003: philosophy has been interpreted as 'anything goes'].

USEFULNESS OF FEEDBACK PASSAGES AFTER DOCUMENT 1
P001: green
P002: yellow
P003: green

Green flag assignment = very useful
Yellow flag assignment = potentially useful

DOCUMENT 2
P004: variety of opinion is necessary for objective knowledge."
P005: the belief that a scientific theory must be under constant scrutiny and that its merit lies only in how well it stands up to rigorous testing.
P006: a theory can be held to be scientific only if it makes predictions that are clearly testable. Critics of this belief acknowledge the strict logic of this process,
P007: Philosophers and historians such as Thomas Kuhn and Paul Feyerabend have attempted to use the history of science

ANSWER CONSTRUCTION:
Popper is a philosopher. Feyerabend [P001: rejected the attempt by certain philosophers to find a methodology applicable to all scientific research.] Feyerabend's [P003: philosophy has been interpreted as 'anything goes']. A positivist is not strict logic. A [P004: variety of opinion is necessary for objective knowledge. A rationalist believes that [P006: a theory can be held to be scientific only if it makes predictions that are clearly testable.] Also that [P005: the belief that a scientific theory must be under constant scrutiny and that its merit lies only in how well it stands up to rigorous testing.]

USEFULNESS OF FEEDBACK PASSAGES AFTER DOCUMENT 2
P001: green
P002: red
P003: green
P004: yellow
P005: green
P006: green
P007: yellow

Red flag assignment = not useful

DOCUMENT 3
P008: observation cannot give explanations based on logic
P009: One controversial point of view is the replacement of scientific realism with scientific relativism

ANSWER CONSTRUCTION:
A rationalist believes that [P006: a theory can be held to be scientific only if it makes predictions that are clearly testable.] Also that [P005: the belief that a scientific theory must be under constant scrutiny and that its merit lies only in how well it stands up to rigorous testing.] Feyerabend [P001: rejected the attempt by certain philosophers to find a methodology applicable to all scientific research.]. [P009: One controversial point of view is the replacement of scientific realism with scientific relativism]. [P002: scientific progress is claimed to be the result of a range of sociological factors working to promote politically convenient notions of how nature operates.] Feyerabend's [P003: philosophy has been interpreted as 'anything goes']. A positivist is not strict logic. [P008: observation cannot give explanations based on logic. Feyerabend is clearly not a rationalist.

USEFULNESS OF FEEDBACK PASSAGES AFTER DOCUMENT 3
P001: green
P002: green
P003: green
P004: red
P005: green
P006: green
P007: red
P008: green
P009: green

DOCUMENT 4
P010: In the 20th century Karl Popper has described scientific method as a rigorous experimental testing of a scientist's ideas or hypotheses
P011: An adequate understanding of the functions of language and of the various types of meaning is another of the fundamentally important contributions of the Logical Positivists.
P012: the merit of a scientific hypothesis lies in its ability to make testable predictions.

ANSWER CONSTRUCTION:
[P012: the merit of a scientific hypothesis lies in its ability to make testable predictions.] A rationalist believes that [P006: a theory can be held to be scientific only if it makes predictions that are clearly testable.] Also that [P005: the belief that a scientific theory must be under constant scrutiny and that its merit lies only in how well it stands up to rigorous testing.] Feyerabend is clearly not a rationalist as he believes [P008: observation cannot give explanations based on logic.] His reasoning for this is that [P002: scientific progress is ~~claimed to be~~ the result of a range of sociological factors working to promote politically convenient notions of how nature operates.]

USEFULNESS OF FEEDBACK PASSAGES AFTER DOCUMENT 4
P001: yellow
P002: green
P003: yellow
P004: red
P005: green
P006: green
P007: red
P008: green
P009: yellow
P010: yellow

P011: red
P012: green

DOCUMENT 5
P013: Popper's principal contribution to the philosophy of science rests on his rejection of the inductive method in the empirical sciences.
P014: According to this traditional view, a scientific hypothesis may be tested and verified by obtaining the repeated outcome of substantiating observations.

ANSWER CONSTRUCTION:
[P012: the merit of a scientific hypothesis lies in its ability to make testable predictions.] A rationalist believes that [P006: a theory can be held to be scientific only if it makes predictions that are clearly testable.] Also that [P005: the belief that a scientific theory must be under constant scrutiny and that its merit lies only in how well it stands up to rigorous testing.] [P013: Popper's principal contribution to the philosophy of science rests on his rejection of the inductive method in the empirical sciences.] Feyerabend is clearly not a rationalist as he believes [P008: observation cannot give explanations based on logic.] His reasoning for this is that [P002: scientific progress is ~~claimed to be~~ the result of a range of sociological factors working to promote politically convenient notions of how nature operates.]

USEFULNESS OF FEEDBACK PASSAGES AFTER DOCUMENT 5
P001: red
P002: green
P003: red
P004: red
P005: green
P006: green
P007: red
P008: green
P009: yellow
P010: red
P011: red
P012: green
P013: green
P014: yellow

DOCUMENT 6
P015: a scientist seeks to discover an observed exception to his postulated rule.

ANSWER CONSTRUCTION:
A rationalist holds [P005: the belief that a scientific theory must be under constant scrutiny and that its merit lies only in how well it stands up to rigorous testing.] [P013: Popper's principal contribution to the philosophy of science rests on his rejection of the inductive method in the empirical sciences.] Popper is a rationalist as he believes that [P014: ~~According to this traditional view,~~ a scientific hypothesis may be tested and verified by obtaining the repeated outcome of substantiating observations.] Feyerabend is clearly not a rationalist as he believes [P008: observation cannot give explanations based on logic.] His reasoning for this is that [P002: scientific progress is ~~claimed to be~~ the result of a range of sociological factors working to promote politically convenient notions of how nature operates.]

USEFULNESS OF FEEDBACK PASSAGES AFTER DOCUMENT 6
P001: red
P002: green
P003: red
P004: red
P005: green
P006: yellow
P007: red
P008: green
P009: red
P010: red
P011: red
P012: red
P013: green
P014: green
P015: yellow

This log tracks the use of feedback passages during the answer construction process. The assignment of flags enables the usefulness of feedback passages to be evaluated. The following variables are also captured in an attempt to better understand the context in which a decision to submit a passage was initially made (see Section 9.2)

CONTEXTUAL UNDERSTANDING
P001: Y
P002: N
P003: Y
P004: N
P005: Y
P006: N
P007: N
P008: Y
P009: Y
P010: Y
P011: Y
P012: N
P013: Y
P014: Y
P015: N

Y=Yes
N=No

INFORMATION NEED ADDRESSED
P001: Y
P002: Y
P003: Y
P004: N
P005: Y
P006: N
P007: N
P008: Y
P009: N
P010: Y
P011: Y
P012: Y
P013: Y
P014: Y
P015: Y

End of Log

APPENDIX F

Multiple Regression:

DERIVING A MODEL OF UNCERTAINTY

INITIAL USEFULNESS JUDGEMENT ASSIGNMENT

Model 1: Number of passages identified by a participant as being not useful.
Model 2: Number of passages identified as being potentially useful + Model 1

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	2.55	.495	9.797	.011
2	2.20	.165	4.379	.066

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.439	1.024		7.267	.000
	Number of passages identified by a participant as being not useful	2.173	.694	.703	3.130	.011
2	(Constant)	10.982	1.910		5.748	.000
	Number of passages identified by a participant as being not useful	1.699	.642	.550	2.649	.027
	Number of passages identified as being potentially useful	-6.114	2.922	-.435	-2.093	.066

a. Dependent Variable: Initial relevance judgement assigned to document (self-reported)

TECHNICAL / AFFECTIVE UNCERTAINTY

Model 1: Indicative or invitational mood.
Model 2: Extent an information need has been addressed + Model 1
Model 3: Conceptual / cognitive state uncertainty + Model 2

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	1.79	.699	23.267	.001
2	1.18	.183	13.990	.005
3	.71	.080	16.715	.003

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.904	1.223		3.193	.010
	Indicative or invitational mood	10.692	2.217	.836	4.824	.001
2	(Constant)	9.560	1.714		5.578	.000
	Indicative or invitational mood	7.120	1.746	.557	4.077	.003
	Extent an information need has been addressed	-14.512	3.880	-.511	-3.740	.005
3	(Constant)	8.790	1.051		8.362	.000
	Indicative or invitational mood	3.336	1.403	.261	2.378	.045
	Extent an information need has been addressed	-14.308	2.342	-.504	-6.110	.000
	Conceptual / cognitive state uncertainty	.277	.068	.412	4.088	.003

a. Dependent Variable: Domain knowledge (self-reported)

DERIVING A MODEL OF TASK COMPLEXITY

NUMBER OF ANSWER CONSTRUCTION INTERACTIONS

Model 1: Incongruities identified.
Model 2: Contextual understanding + Model 1
Model 3: Extent an information need has been addressed + Model 2
Model 4: Number of passages highlighted + Model 3

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	2.21	.791	37.888	.000
2	1.83	.081	5.694	.041
3	1.55	.046	4.455	.068
4	1.29	.033	4.588	.069

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	19.760	1.798		10.991	.000
	Incongruities identified	-25.074	4.074	-.889	-6.155	.000
2	(Constant)	14.920	2.513		5.938	.000
	Incongruities identified	-19.907	3.998	-.706	-4.979	.001
	Contextual understanding	6.500	2.724	.338	2.386	.041
3	(Constant)	11.307	2.737		4.131	.003
	Incongruities identified	-16.333	3.797	-.579	-4.302	.003
	Contextual understanding	5.551	2.359	.289	2.353	.046
	Extent an information need has been addressed	5.067	2.401	.267	2.111	.068
4	(Constant)	6.525	3.187		2.047	.080
	Incongruities identified	-15.772	3.166	-.560	-4.983	.002
	Contextual understanding	7.714	2.205	.402	3.499	.010
	Extent an information need has been addressed	6.873	2.166	.362	3.174	.016
	Number of passages highlighted	10.303	4.810	.244	2.142	.069

a. Dependent Variable: Number of answer construction interactions

INTRINSIC COGNITIVE LOAD

MODEL 1: KNOWLEDGE OBTAINED FROM PREVIOUS SEARCH INTERACTIONS

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	1.25	.740	28.437	.001

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.249	.846		-1.476	.171
	Knowledge obtained from previous search interactions	.433	.081	.860	5.333	.000

a. Dependent Variable: Intrinsic cognitive load (computation)

EXTRANEOUS COGNITIVE LOAD

MODEL 1: TASK COMPLEXITY

Model	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	1.36	.754	30.638	.001

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.427	.811		-1.760	.109
	Task complexity	9.424	1.703	.868	5.535	.000

a. Dependent Variable: Extraneous cognitive load (computation)