A FRAMEWORK FOR THE DESIGN OF USABLE ELECTRONIC TEXT

BY

ANDREW PATRICK DILLON, B.A., M.A.(CORK)

A DOCTORAL THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY OF THE LOUGHBOROUGH UNIVERSITY OF TECHNOLOGY

APRIL 1991

A FRAMEWORK FOR THE DESIGN OF USABLE ELECTRONIC TEXT.

This thesis examines the human issues underlying the design and usability of electronic text systems. In so doing it develops a framework for the conceptualisation of these issues that aims to guide designers of electronic texts in their attempts to produce usable systems.

The thesis commences with a review of the traditional human factors literature on electronic text according to three basic themes: its concern with perceptual, manipulatory and structural issues. From this examination it is concluded that shortcomings in translating this work into design result from the adoption of overly narrow uni-disciplinary views of reading taken from cognitive psychology and information science which are inappropriate to serve the needs of electronic text designers.

In an attempt to provide a more relevant description of the reading process a series of studies examining readers and their views as well as uses of texts is reported. In the first, a repertory grid based investigation revealed that all texts can be described in reader-relvant terms according to three criteria: why a text is read, what a text contains and how it is read. These criteria then form the basis of two investigations of reader-text interaction using academic journals and user manuals.

The results of these studies highlighted the need to consider readers' models of a document's structure in discussing text usability. Subsequent experimental work on readers' models of academic articles demonstrated not only that such models are important aspects of reader-text interaction but that data of this form could usefully be employed in the design of an electronic text system.

The proposed framework provides a broad, qualitative model of the important issues for designers to consider when developing a product. It consists of four interactive elements that focus attention on aspects of reading that have been identified as central to usability. Simple tests of the utility and validity of the framework are reported and it is shown that the framework both supports reasoned analysis and subsequent prediction of reader behaviour as well as providing a parsimonious account of their verbal utterances while reading. The thesis concludes with an analysis of the likely uses of such a framework and the potential for electronic text systems in an increasingly information-hungry world.

Acknowledgments

Writing a thesis could be described as a labour of love but by the time final drafts are being examined it is merely a labour. The lingering memory is primarily one of lost weekends and endless evenings sitting alone in front of a word processor and a mess of notes. However, the trials and tribulations have been lessened by a few who out of interest, kindness or foolishness (or a wonderful mixture of all three) contributed to these efforts in varying ways. Their reward is only a mention here and their justified smugness at having greased the wheels of science! The honours list follows.

Dr. Cliff McKnight, a supervisor and co-researcher, who, despite wondering what I was raving on about half the time, enabled the whole process, offered soothing words over liquid lunches and always bought his round;

Prof. Ken Eason, Director of Research who untied the administrative knots, spoke words of encouragement and wisdom and whose easy manner betrayed the hard-earned knowledge of exactly what it is like to be a part-time student with a full-time job;

John Richardson, co-researcher whose sound judgement was always a good testing ground for ideas and who contributed to the work reported here in myriad ways;

David Davies, whose interest and enthusiasm for the thesis started the ball rolling and added momentum at times of high viscosity;

The HUSAT Research Institute, which supported the whole process;

The British Library Research and Development Dept. and OCLC, Ohio, USA, both of which indirectly funded the process by financing three projects at HUSAT on issues of electronic text that provided much of the subject matter of this thesis;

Marian Sweeney, Oh what can I say?

Finally, to the subjects who gave so willingly (well, sometimes) of their time and intellectual efforts.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

THE READING PROCESS WITH RESPECT TO ELECTRONIC TEXT

1.1 Introduction

Information technology continues to develop swiftly and while it is typical to discuss this in terms of falling hardware costs and technological advances (e.g., Osborne 1979), it is for the psychologist at least, more interesting to observe such developments in terms of their influence on human activities. While theorists talk of the information age, we are in fact creating an information world where microprocessors interface between us and innumerable as well as previously unimaginable activities.

The present work is concerned with one such development, the use of information technology to support the activity known as "reading" and in so doing, to challenge the supremacy of paper as the most suitable medium of text presentation. This area is receiving a lot of attention currently as hypertext gives new substance to old ideas but such attention is directed more at developing the technology than considering how and why it might be useful. This thesis is concerned with the reader.

1.2 The emergence of electronic text

For a medium that is so new it is perhaps surprising that a history of electronic text can be even considered never mind described. However, the idea of using the electronic medium to support reading can be traced back several decades and no self-respecting writer on the subject of hypertext ever fails to mention such visionary thinkers as Bush (1945), Engelbart (1963) or Nelson (1987)[1] who in their own way advanced (and in some cases continue to do so) the concept of access to a world of knowledge through information technology. These thinkers paved the intellectual path to hypertext and its underlying philosophy that humans should be able, from their desktop, to locate, retrieve and use easily the store of human knowledge that lies in books, journals and associated materials in libraries the world over.

Despite its ancestry or philosophy, electronic text has had to wait for a technology to develop before such fantastic ideas could be embodied. The computer is that technology

[1] The comparative recency of this reference stems from the fact that much of Nelson's work is self-published and/or distributed. His ideas actually gained currency far earlier than this date suggests.

and only comparatively recent developments in microelectronics have enabled the concept of electronic text to be seen and not just heard. Feldman (1990) points out that despite the advocates of previous decades it is only in the 1980s that electronic text could really be said to have arrived. Prior to that it was conceived, talked about, and its potential imagined, but it did not truly exist.

The precision of the timeline is debatable but it is not difficult to see why the personal computer boom of the 1980s coupled with developments in digital information storage and presentation have made electronic text both feasible and culturally acceptable. Both these aspects are necessary for electronic text to succeed. It is not enough that it can now be done, that electronic text can, for example, reduce the 20-volume Grolier Encyclopedia to a single compact disc (and still leave more than half of the disk free), but the world needs to be ready for electronic text. Readers must appreciate its relevance, its potential advantages and more importantly, they must want electronic text if it is to succeed.

While the information culture is emerging, the acceptance of electronic text currently lags behind its technical feasibility. With the exception of a small number of researchers, designers and keen amateurs, the idea of reading lengthy texts in their electronic as opposed to paper form tends to be viewed negatively rather than embraced wholeheartedly. It will take time and effort to identify the optimum form for electronic text, current designers lack the guidelines and experiences that have evolved with paper text and the present thesis seeks to contribute to this effort. These are early days for the new medium (even if it is possible to distinguish between generations of electronic text) and one should avoid seeing electronic text as a competitor to paper in some form of "either-or" challenge for supremacy. It is not inevitable that electronic text will replace paper as some writers have suggested (e.g., Jonassen 1982) but it might displace it as more and more human activities become mediated by information technology. This should not be allowed to happen by accident though, as a side effect of increased computerisation, we must seek to actively influence the process for the better so that the positive aspects of electronic text are accentuated. To achieve this psychologists and other human scientists must influence designers. The history of electronic text is still being written.

1.3 The aims of the thesis

The major aim of the present work is to examine and describe the reading process from a perspective that sheds light on the potential for information technology to support that

process. Current research suggests that paper is by far the preferred medium for reading though there is less consensus on why this is, in fact, the case. It is clear that simply transferring paper formats to the electronic medium is insufficient and often detrimental to use. Therefore clarification of the future role for electronic text in our information world would seem to be an issue worthy of investigation. Tackling it as a psychologist, the author is far less concerned with technical feasibility i.e., can it be built? (although at times this rears its head) than with how the human cognitive system places constraints on, or provides clues to, the usability of current and future technologies.

Traditionally, human factors has offered itself to design engineers as an evaluative discipline, equipped with the tools to assess the performance of human operators with developed systems. In recent years, as a result of more rapid design cycles and increased competition amongst developers, a need for earlier inputs to the product life cycle has arisen. Such inputs, in the form of models, guidelines, checklists and design tools attempt to package ergonomic knowledge in a form suitable for engineers to consume and apply. This has not proved an easy task and there are many in the human factors discipline uncomfortable with this role.

This thesis will not directly tackle that issue but is aware of its existence as a yardstick against which the value of current human factors work is often measured. Consequently a second aim is to develop a framework for considering user issues that is applicable to the earliest stages of electronic text design. The emphasis throughout the work is therefore less on empirical investigations of various user interface variables (though these are present) and more on identifying the crucial psychological factors underlying reading through knowledge elicitation techniques and observation of usage patterns with a view to forming these into a conceptual framework that can be presented as a "package" of human factors knowledge relevant to design.

1.4 The scope of the thesis

In simple terms, this work is concerned with the human as reader i.e., user of textual information. However, its remit is broad by comparison to much of the theoretical psychological work in this area which tends to define reading narrowly as the transformation of visual images to perceived words or the extraction of meaning from structured prose.[2] Rather it covers the range of issues involved in using such material,

[2] Such positions, though extremist, are both tenable and frequently published. Indeed much

from identifying a need for it, locating and selecting it, manipulating it and ultimately processing it. Therefore, while its interests are primarily psychological, the consideration of alternative perspectives from disciplines such as information science, computer science and typography are both necessary and insightful.

In the present context therefore "reading" implies situations where the human will engage the medium to perform any of the range of activities typically afforded this descriptor in common parlance. Thus it covers a variety of scenarios ranging from proof-reading to examining contents but omits those that have reading as a component but necessarily secondary part, such as text-editing. Furthermore, included under this term are the associated activities of location and manipulation of textual information that invariably precede and are concurrent with these tasks in the real world.[3]

By text is meant any document, with or without graphics, that can be presented to a reader as an information source. Thus it includes those documents that we are typically exposed to in everyday life such as newspapers, books, magazines, technical manuals, letters and so forth, as well as less traditional "texts" such as electronic documents and databases. Though termed text, this descriptor might include those documents that have a large graphical content (such as catalogues) but not those that are primarily graphical such that they relegate alphanumeric text strings to secondary importance (for example, maps).

The term electronic text is used by means of contrast with paper documentation i.e., it is any text that is presented on a computer screen.[4] It includes hypertext and non-hypertext. Like its paper equivalent it may contain graphics. However it does not cover the term hypermedia which is often mistakenly assumed to be hypertext with graphics.[5] Generically, the term information space is used to cover all published materials: text,

---

of the work in experimental psychology on reading assumes one or other interpretation (see e.g. Just and Carpenter's (1980) model of reading which includes comprehension and Crowder's (1982) description of where reading begins and ends which explicitly excludes it).

[3] How far one extends this is a matter of common-sense. Obviously walking into a library necessarily precedes the act of reading there but should not be considered part of the act itself. However within the broad task scenario, searching for a book or browsing the spines of numerous journals in order to locate a specific edition are part of reading in this sense.

[4] For the purposes of this thesis the terms electronic and screen-presented text are used synonymously and imply presentation via computer screens. They do not refer to any other form of screen presented text such as microfiche, microfilm or slides which involve magnification and projection rather than electronic processing.

[5] These terms are often used synonymously in the literature which is erroneous. Hypermedia implies the use of different media to present a range of information such as text, sound and animation, not just the marriage of text and graphics on a single medium. Hypermedia information sources have no paper equivalent.

hypertext and hypermedia. Where it is employed in this thesis its meaning is implied by the context of use unless otherwise stated.

Obviously it is impossible in the present situation to cover all manifestations of reading texts as the terms are defined here and indeed the thesis concentrates primarily (though not completely) on academic literature such as journal articles for its empirical investigations. However, academic articles are lengthy texts which, it will be shown, are read in a variety of ways that extend their comparability with other document forms. Furthermore, the use of both software manuals and a booklet in two of the studies reported here broaden the coverage of the thesis. Thus the issues raised and concepts presented in the final framework are intended to be generic and applicable to most text forms and reading situations in as much as electronic media might influence their interaction.

1.5 A note on methods

This work, by choice, avoided many of the issues of learning to use innovative technology which some would see as a natural role for a human factors study. It is not that such research is seen as irrelevant but that the author believes that well-designed systems should start from a premise of supporting certain tasks rather than worrying prematurely about ease of learning. In this application domain design is necessarily speculative, there are few if any rules or established systems to react to or design against.[6] Consequently, the author sees the role of psychology in this area as a dual one of guidance and suggestion, using its knowledge of human cognition to constrain the number of potential design options while informing speculation on how humans might like things to be. Such work necessarily precedes learnability research. Put simply, electronic text is at a stage of development where its potential use rather than its ease of use is of more concern.

The stated aims and approach of the thesis have dictated the methods employed. This thesis is an applied work, a study of human factors issues carried out during the development and evaluation of real products. In order to identify how electronic text systems are designed and the best role for human factors knowledge in this process it is necessary to involve oneself in the process, to be part of a design team, to develop electronic texts and assess the consequences of one's work. Only in this way can one really appreciate what is needed, what questions arise, what type of human factors input is useful and what are the limitations of the discipline's (and one's own) knowledge. Theorising from without may have proved intellectually stimulating but would have been insufficient. To paraphrase Card

[6]Contrast this with designing a new text editor where not only does a large body of knowledge on how users perform such tasks exist but designers can examine numerous existing products to inform their own design.

et al. (1983), "design is where the action is" in HCI.

On the face of it, the human scientist would appear not particularly well armed for action of this kind. The traditional strengths of psychology and ergonomics lie in designing and conducting formal experiments, planning work in detail in advance of carrying it out, controlling for all undesirable sources of variance and reporting the results in conventional academic form. As a result of such an approach a substantial literature has emerged on the usability of various interface features or the significant problems associated with certain products. Essential as such work is in building up the bedrock of empirical knowledge, on its own it cannot provide the answers to the questions posed here.

Examining the issue more deeply however, one might come to see that the human scientist is the ideal person to become involved in the design of interactive products. Equipped with knowledge of human behaviour and dispositions, skilled in the consideration of how certain design features influence performance, the human scientist can make the distinction between popular conceptions of users based on opinion and myth and accurate models based on reasoned argument and psychological findings. He should be able to distinguish between occasions when approximate answers will suffice and when only formal experimental evaluation will provide answers. Most importantly, he should be able to identify gaps in the knowledge base of design that only the human sciences can fill or hope to fill. In short, the human scientist may be seen as the only suitable candidate for the job.

This is the philosophy of the present work. Involvement has been achieved by working at the HUSAT Research Institute in Loughborough,[7] the largest university-based research and consultancy institute in Europe dedicated to the application of the tools and techniques of the human sciences in technology design. The author became a member of a research team in 1987 investigating electronic text design on three projects. In the first, Project Quartet (Tuck *et al.*, 1990) the author was a member of a three-man team of psychologists at HUSAT who worked with other research teams at three universities consisting, in the main, of computer scientists. The goal of this work was to investigate the impact of information technology on scholarly communication. In the other two projects (Project OCLC [McKnight *et al.*, 1988] and Project CHIRO [McKnight *et al.*, 1990c]), the author was a member of the same team at HUSAT investigating both the interface issues associated with access, delivery and usage of lengthy electronic texts from remote sources, and the use of hypertext as a collaborative work environment for scientists. These projects, coupled with a variety of short-term consultancies on human factors for numerous industrial software companies and departments, were, and continue to be, the testing

[7] HUSAT stands for Human Sciences and Advanced Technology.

grounds for the ideas outlined in this work.

The nature of the work and the impact of the author's findings on real-world applications means that most of the experiments and studies reported here are not laboratory exercises isolated from practical concerns, but investigations carried out during design processes to provide answers to genuine questions, to resolve design issues or to test specific design instantiations. It is an example of the psychologist as applied scientist, part designer, part team member and part user, all the while monitoring his and his colleagues' own work in a meta-analytic fashion. It is the author's contention that such a process is not only a worthwhile method of research but the only sure way for suitable knowledge of this field to be gained.

Generally, the techniques and methods employed in this thesis varied from the formal experimental to the exploratory. More specifically they ranged from controlled laboratory experiments to interviews with users, with a selection of others in between. The use of a method was determined by the type of information sought – what needed to be known led to the choice of investigative methodology. Expertise in a technique was never considered sufficient justification for its employment. For example, at the outset it became clear that information on how readers view texts and their interactions with them was in short supply in the literature. This gap in the knowledge base is in part due to the inherent difficulties in capturing such information in a valid form. Experimental techniques are impracticable in such situations and reliable questionnaires on such matters do not yet exist. In order to overcome this, information was gathered employing a mix of knowledge elicitation techniques from the "harder"[8] or more objective such as repertory grid analysis, to the "softer" or more subjective ones such as interviewing until satisfactory answers were obtained. In this the author took inspiration from Binder (1964), a psychologist and statistician who wrote:

> "We must use all available weapons of attack, face our problems realistically and not retreat to the land of fashionable sterility, learn to sweat over our data with an admixture of judgement and intuitive rumination, and accept the usefulness of particular data even when the level of analysis for them is markedly below that for other data in the empirical area" (p. 294).

Later work, particularly the testing of predictions derived from the framework, allowed a more "rigourous" approach to data collection and several experimental studies are included. Three of these were carried out as part of a real-world design process i.e., they were to inform the design of a hypertext database of journal articles undergoing development at

[8]"Harder" in the comparative sense only amongst knowledge elicitation techniques-i.e., more objective than interviews.

HUSAT. As such they are both academic exercises and tests of how human factors might fit into system development. The latter requirement placed constraints on resources allocated to these investigations. While it might have been academically more pure to increase sample sizes and tighten controls, the demands for answers and the window of opportunity for the input necessitated a slightly "dirtier" approach. The results however are valid and testable. The resulting database is the embodiment of the approach and initial feedback from users has been positive (see e.g., McKnight *et al.*, 1990b).

As a result of working in a team it is impossible to isolate whole aspects of work for oneself. Ideas are exchanged, other's thoughts and opinions influence one's own views and responsibilities for data gathering are often shared. The work reported here is in parts no exception and while every attempt has been made to report only work that the author was solely responsible for, the involvement of other people in certain studies has been unavoidable. To clarify the situation, each chapter covering experimental or other data gathering work contains footnotes or sections explicitly outlining the work of the author and the exact nature of the involvement of other people. It can be stated clearly though that the interpretation of all data presented in this thesis is the author's work. Furthermore, the final framework, resting as it does on an interpretation of what is required and how design proceeds in the electronic text domain is entirely the present author's work.

1.6 Outline of the thesis

The thesis commences with a thorough review of the literature on reading from paper and screens. This review is divided into three major parts. The first describes the reported differences between reading from paper and from screens. It is shown that five broad outcome and three process differences have been observed.

Part two reviews the analyses of these differences that have resulted. These analyses address ergonomic comparisons between reading from paper and screen that have emerged in the last fifteen years. These can be seen as an attempt to identify the basic differences between the media and subsequently isolate the crucial variables in terms of three levels: perceptual, motor and cognitive isues. Shortcomings of this work in terms of methodological weakness and ecological validity are highlighted. The third part of the literature review ends with a categorisation of the crucial factors for developing electronic texts.

Chapter three concentrates on the value of the human factors literature to designers, highlighting its underlying acceptance of narrow uni-disciplinary views of the reading

process. The problems of applying findings from this work to the design of an electronic text system are highlighted by reference to a case study carried out by the author.

In order to overcome the shortcomings of previous work it was felt that an understanding of readers' views of text usage would be useful. To this end a repertory grid analysis of six individuals' perceptions of a range of text types was carried out and is reported in chapter four. The results of this classification study provided a stimulus to more detailed interviews with users of two demonstrably distinct text types: academic journals and software manuals which are reported in chapter five. In combination these data shed light on what readers see as important in text usage and how electronic versions might be designed to convey benefits.

On the basis of this work it was felt that the concept of readers' mental models of the information structure inherent in texts required closer examination. Chapter six considers further the literature on this aspect of reading and reports on two studies carried out by the author to demonstrate its relevance to electronic text design in general and to the design of a hypertext database at HUSAT in particular. The results indicate that for academic articles this model is strong and facilitates prediction and location of material within the body of the text.

Following this work the framework is proposed in chapter seven. This represents the psychological factors involved in using a text and suggests the variables to consider in designing an electronic text. The framework consists of four interactive elements which exert control over the reader's attention at various stages of the reading process. The interactions between these elements are described and the role of such frameworks in human factors work is assessed.

This framework is then used to derive predictive models of reader performance on certain tasks, and these are tested in two studies using hypertext and paper. Verbal protocols from one study show strong support for the elements described in the framework. Performance on the tasks in both studies generally support the predictions from the models. It is concluded that the framework is both a valid description of the reading process for considering electronic text and can usefully support human factors predictions at the early stages of design.

The final chapter reviews the work of the thesis in the light of its stated aims and highlights areas for future research. A sequence of human factors inputs to the design stage which should aid usability of the resultant application is made explicit in this chapter. The final

section of the thesis indicates the realistic prospects for electronic text in the information world that is being created.

CHAPTER 2

THE READING PROCESS AND ELECTRONIC TEXT: A REVIEW OF ISSUES

2.1 Introduction

In simple terms, there exist two schools of thought on the subject of electronic texts. The first holds that paper is far superior and will never be replaced by screens. Such arguments are frequently supported by reference either to the type of reading scenarios that would currently prove difficult if not impossible to support acceptably with electronic text, e.g., reading a newspaper on the beach or a magazine in bed, or the unique tactile qualities of paper. The latter aspect is summed up neatly in Garland's (1982) comment that electronic text may have potential uses:

> "but a book is a book is a *book*. A reassuring, feel-the-weight, take-your-own-time kind of thing..." (cited in Whaller 1987, p. 261).

The second argument favours the use of electronic text, citing ease of storage and retrieval, flexibility of structure and saving of natural resources as major incentives. According to this perspective, electronic text will soon replace paper and in a short time (usually ten years hence) we shall all be reading from screens as a matter of habit. In the words of its greatest proponent, Ted Nelson (1987):

> "the question is not can we do everything on screens, but when will we, how will we and how can we make it great? This is an article of faith - its simple obviousness defies argument."[1]

Such extremist positions show no signs of abating though it is becoming clear to many researchers in the domain that neither is particularly accurate. Reading from screens *is* different from paper and there are many scenarios such as those cited that current technology would not support well, if at all. However, technology is developing and electronic text of the future is unlikely to be handicapped by limitations in screen image and portability that currently seem major obstacles. As Licklider pointed out when considering the application of computers in libraries as early as 1965:

---

[1] As Nelson's book is distributed as a hypertext document there are no page numbers. However, this quote can be located in Chapter 1, An Obvious Vision, under the heading Tomorrow's World of Text on Screen. Such lengthy reference to a specific section highlights a potential problem with hypertext that must be addressed.

> "our thinking and our planning need not be, and indeed should not be, limited by literal interpretation of the existing technology" (p.19).

Even so, paper is an information carrier *par excellence* and possesses an intimacy of interaction that can never be obtained in a medium that by definition imposes a microchip interface between the reader and the text. Furthermore, the millions of books that exist now will not all find their way into electronic form, thus ensuring the existence of paper documentation for many years yet.

The aim of the present review is not to resolve the issue of whether one or other medium will dominate but to examine critically the reported differences between them in terms of use and thereby support reasoned analysis of the paper versus electronic text debate from the perspective of the reader. In so doing it should highlight the crucial issues underlying the usability of a medium and thereby inform the development of the framework for conceptualising the reading process, the ultimate aim of this thesis.

2.2 The outline of the review

The review is divided into three major parts although each of these is sub-divided into several levels. The first part (sections 2.3 to 2.5) describes the reported differences between the media, i.e., what human variables are altered when a reader uses an electronic rather than a paper text. To this end it draws a distinction between outcome and process differences in reading and tries to concentrate only on the available evidence rather than theorists' views. The second major part (sections 2.6 to 2.9) covers the various attempts that have been made to explain or overcome these differences and identifies three levels of explanation covering perceptual, motor and cognitive factors. Again, experimental work is focused on although the views of researchers as to why particular variables or groups of variables are important are also discussed. The third and final major part (section 2.10) tries to draw the various strands together into a cohesive view of what is known about reading electronic text and how progress can best be made.

At the outset it must be stated that, although extensive, drawing any firm conclusions from the literature is difficult. Helander *et al.* (1984) evaluated 82 studies concerning human factors research on reading from VDUs and concluded:

> "Lack of scientific rigour has reduced the value of many of these studies.

> Especially frequent were flaws in experimental design and subject selection, both of which threaten the validity of results. In addition, the choice of experimental settings and dependent and independent variables often made it difficult to generalize the results beyond the conditions of the particular study" (p.55).

Waern and Rollenhagen (1983) point to the frequently narrow scope of experimental designs in such studies. Important factors are either not properly controlled or are simply not reported and most studies use unique procedures and equipment, rendering direct comparison meaningless. The present review is not intended to untangle the methodological knots of other researchers but rather to make sense of the major findings in a general way, indicating where the research needs lie and what issues a framework intended to support the process of electronic text design must include.

A detailed literature already exists on typographical issues related to text presentation on paper (see particularly the work of Tinker 1958 and 1963) and issues such as line spacing and formatting are well researched. This work will not be reviewed here as much of it remains unreplicated on VDUs and evidence suggests that, even when such factors are held constant, reading differences between the two presentation media remain (see for example Creed *et al.,* 1987). This review will concentrate on identifying the nature of any differences that may exist between reading from paper and screens, their possible causes, and under what conditions, if any, they may be resolved.

2.3 Observed differences: outcome versus process measures

Analysing reading is not a simple task and a distinction has been drawn between assessing reading behaviour in terms of outcome and process measures (Schumacher and Waller 1985). Outcome measures concentrate on what the reader gets from the text and considers such variables as amount of information retrieved, accuracy of recall, time taken to read the text and so forth. Process measures are more concerned with how the reader uses a text and include such variables as where the reader looks in the text and how he manipulates it. Obviously the two types of measure are inter-related, outcome will be affected by process and process will often depend on the desired outcome for example.

In the domain of electronic text outcome measures take on a particular relevance as advocates proclaim increased efficiency and improved performance (i.e. outcomes) with computer presented material (aspects of direct concern to human factors

psychologists). It is not surprising therefore to find that the majority of work comparing the two media has concentrated heavily on such differences. With the emergence of hypertext however, navigation has become a major issue and process measures are gaining increased recognition of importance.

In the following sections a summary of the observed differences between the media in terms of outcomes and processes is presented. It should be noted however that process and outcome are not perhaps as discrete as research might suggest. As indicated, from the reader's perspective process leads to output and desired output will influence process. The distinction emphasised here is one of measurement, not one of activity.

2.4 Outcome measures

2.4.1 Speed

By far the most common experimental finding is that reading from screen is significantly slower than reading from paper (Kak, 1981; Muter *et al.*, 1982; Wright and Lickorish, 1983; Gould and Grischkowsky, 1984; Mills and Weldon, 1985). Figures vary according to methods of calculation and experimental design but the evidence suggests a performance deficit of between 20% and 30% when reading from screen.

However, despite the apparent similarity of findings, it is not clear whether the same mechanisms have been responsible for the slower speed in these experiments, given the great disparity in procedures. For example, in the study by Muter *et al.* (1982), subjects read white text on a blue background, with the subject being approximately 2.5 m from the screen. The characters, displayed in teletext format on a television, were approximately 1 cm high, and time to fill the screen was approximately 9 seconds. Even ignoring the unnatural character size and distance from the screen, the authors reported that the experimental room was "well illuminated by an overhead light source", a factor which by virtue of the reflections caused could account for a slow reading speed. Additionally, unless the book used was one of the large format books prepared for the partially sighted, one must assume that the screen text characters were substantially bigger than the printed characters.

In comparison, Gould and Grischkowsky (1984) used greenish text on a dark background. Characters were 3 mm high and subjects could sit at any distance from the screen. They were encouraged to adjust the room lighting level and the luminance and

contrast of the screen for their comfort. Printed text used 4 mm characters and was laid out identically to the screen text. Wright and Lickorish (1983) give no details of text size other than that it was displayed as white characters on a black 12" screen driven by an Apple][microcomputer with lower case facility. This would suggest that it was closer to Gould's text than Muter's text in appearance. Printed texts were photocopies of printouts of the screen displays produced on an Epson MX-80 dot matrix printer, compared with Gould's 10-point monospace Letter Gothic font.

In contrast to these studies, Switchenko (1984), Askwall (1985) and Cushman (1986) found that reading speed was unaffected by the presentation medium. Askwall attributes this difference in findings to the fact that her texts were comparatively short (22 sentences), and the general lack of experimental detail makes alternative interpretations difficult. Although it is reported that a screen size of 24 rows by 40 columns was used, with letter size approximately 0.5 x 0.5 cm and viewing distance of approximately 30-50 cm, no details of screen colour or image polarity and none of the physical attributes of the printed text are given.

Cushman's primary interest was in fatigue but he also measured reading speed and comprehension using 80-minute reading sessions. Negative and positive image VDU and microfiche presentations were used and most of the 76 subjects are described as having had "some previous experience using microfilm readers and VDUs." On the basis of this study Cushman concluded that there was no evidence of a performance deficit for the VDU presentations compared with printed paper.

As this indicates, the evidence surrounding the argument for a speed deficit in reading from VDUs is less than conclusive. A number of intervening variables, such as the size, type and quality of the VDU may have contaminated the results. As will be consistently demonstrated, this criticism applies repeatedly to most of the evidence on reading from screens. However, despite the methodological weaknesses of many of the investigations, evidence continues to mount supporting the case for a general speed decrement. As Gould *et al.* (1987a) noted, many of these experiments are open to interpretation but :

> "the evidence on balance...indicates that the basic finding is robust– people do read more slowly from CRT displays" (p. 269).

2.4.2 Accuracy

Accuracy of reading could refer to any number of everyday activities such as locating information in a text, recalling the content of certain sections and so forth. In experimental investigations of reading from screens the term accuracy has several meanings too, though it most commonly refers to an individual's ability to identify errors in a proofreading exercise. While a number of studies have been carried out which failed to report accuracy differences between VDUs and paper (e.g., Wright and Lickorish, 1983; Gould and Grischkowsky, 1984) recent well controlled experiments by Creed *et al.* (1987) and Wilkinson and Robinshaw (1987) report significantly poorer accuracy for such proofreading tasks on screens.

Since evidence for the effects of presentation media on such accuracy measures invariably emerges from the same investigations which looked at the speed question, the criticisms of procedure and methodology outlined above apply equally here. Furthermore, the measures of accuracy employed also vary. Gould and Grischkowsky (1984) required subjects to identify misspellings of four types: letter omissions, substitutions, transpositions and additions, randomly inserted at a rate of one per 150 words. Wilkinson and Robinshaw (1987) argue that such a task hardly equates to true proof reading but is merely identification of spelling mistakes. In their study they tried to avoid spelling or contextual mistakes and used errors of five types: missing or additional spaces, missing or additional letters, double or triple reversions, misfits or inappropriate characters, and missing or inappropriate capitals. It is not always clear why some of these error types are not spelling or contextual mistakes but Wilkinson and Robinshaw suggest their approach is more relevant to the task demands of proofreading than Gould and Grischkowsky's.

However Creed *et al.* (1987) distinguished between visually similar errors (e.g., "e" replaced by "c"), visually dissimilar errors (e.g., "e" replaced by "w") and syntactic errors (e.g., "gave" replaced by "given"). They argue that visually similar and dissimilar errors require visual discrimination for identification while syntactic errors rely on knowledge of the grammatical correctness of the passage for detection and are therefore more cognitively demanding. This error classification was developed in response to what they see as the shortcomings of the more typical accuracy measures which provide only gross information concerning the factors affecting accurate performance. Their findings indicate that visually dissimilar errors are significantly easier to locate than either visually similar or syntactic errors (in other words, the more demanding the task, the less accurate is reading from screens).

In a widely reported study Egan *et al.* (1989) compared students' performance on a set of tasks involving a statistics text presented on paper or screen. Students used either the standard textbook or a hypertext version run on SuperBook, a structured browsing system, to search for specific information in the text and write essays with the text open.[2] Incidental learning and subjective ratings were also assessed. The search tasks provide an alternative to, and more realistic measure of, reading accuracy than identifying spelling errors.

The authors report that subjects using the hypertext performed significantly more accurately than those using the paper text. However a closer look at the experiment is revealing. With respect to the search tasks, the questions posed were varied so that their wording mentioned terms contained in the body of the text, in the headings, in both of these or neither. Not surprisingly the largest advantage to electronic text was observed where the target information was only mentioned in the body of text (i.e. there were no headings referring to it). Here it is hardly surprising that the search facility of the computer outperformed humans (though this does highlight one of the major advantages of electronic text that should be exploited). When the task was less biased against the paper condition e.g. searching for information for which there are headings, no significant difference was observed. Interestingly the poorest performance of all was for SuperBook users searching for information when the question did not contain specific references to words used anywhere in the text. In the absence of suitable search parameters or look-up terms hypertext suddenly seemed less usable.

Regardless of the interpretation that is put on the results of any one of these studies, the fact remains that investigations of reading accuracy from screen and paper take a variety of measures as indices of performance. Therefore two studies, both purporting to investigate reading accuracy may not necessarily measure the same events. It would seem that for routine spelling checks reading from screens is not less accurate than reading from paper. However, a performance deficit does seem to occur for more visually or cognitively demanding tasks.

2.4.3 Fatigue

The proliferation of information technology has traditionally brought with it fears of

---

[2] There is some debate over whether or not SuperBook is really hypertext which stems from the various defining characteristics of hypertext. The developers point of view is that it doesn't really matter what it is called, SuperBook is an advanced text presentation system supporting flexible access and therefore will be considered hypertext for the purposes of this thesis.

harmful or negative side-effects for users who spend a lot of time in front of a VDU (see for example Pearce, 1984). In the area of screen reading this has manifested itself in speculation of increased visual fatigue or eyestrain when reading from screens as opposed to paper.

In the Muter *et al.* study (op cit) subjects were requested to complete a rating scale on a number of measures of discomfort including fatigue and eyestrain both before and after exposure to the task. There were no significant differences reported on any of these scales either as a result of condition or time. Similarly Gould and Grischkowsky (1984) obtained responses to a 16-item "Feelings Questionnaire" after each of six 45-minute work periods. This questionnaire required subjects to rate their fatigue, levels of tension, mental stress and so forth. Furthermore various visual measurements such as flicker and contrast sensitivity, visual acuity and phoria, were taken at the beginning of the day and after each work period. Neither questionnaire responses nor visual measures showed a significant effect for presentation medium. These results led the authors to conclude that good-quality screens in themselves do not produce fatiguing effects, citing Starr (1984) and Sauter *et al* (1983) as supporting evidence.

In a more specific investigation of fatigue Cushman (1986) investigated reading from microfiche as well as paper and VDUs with positive and negative image[3]. He distinguished between visual and general fatigue, assessing the former with the Visual Fatigue Graphic Rating Scale (VFGRS) (a scale devised by Cushman himself) which subjects use to rate their ocular discomfort, and the latter with the Feeling-Tone Checklist (FTC, Pearson and Byars, 1956). With respect to the VDU conditions, the VFGRS was administered before the session and after 15, 30, 45 and 60 minutes as well as at the end of the trial at 80 minutes. The FTC was completed before and after the session. The results indicated that reading from positive presentation VDUs (dark characters on light background) was more fatiguing than paper and leads to greater ocular discomfort than reading from negative presentation VDUs.

Cushman explained the apparent conflict of these results with the established literature, which reports no such differences, in terms of the refresh rate of the VDUs employed (60 Hz) which may not have been enough to completely eliminate flicker in the case of positive presentation, a suspected cause of visual fatigue. Wilkinson and Robinshaw (1987) also reported significantly higher fatigue for VDU reading and while their equipment may also have influenced the finding they dismiss this as a reasonable

[3] The terms positive and negative presentation refer to the polarity of the image and background. Positive presentation is used here to mean dark text on a light background, negative presentation therefore refers to light text on a dark background.

explanation on the grounds that no subject reported lack of clarity or flicker and their monitor was typical of the type of VDU that users find themselves reading from. They suggest that Gould and Grischkowsky's (1984) equipment was "too good to show any disadvantage" and that their method of measuring fatigue was artificial. By gathering information after a task and across a working day Gould and Grischkowsky missed the effects of fatigue within a task session and allowed time of day effects to contaminate the results. Wilkinson and Robinshaw liken the proofreading task used in these studies to vigilance performance and argued that fatigue is more likely to occur within the single work period where there are no rest pauses allowing recovery. Their results showed a performance decrement across the 50-minute task employed leading them to conclude that reading from typical VDUs at least for periods longer than 10 minutes is likely to lead to greater fatigue.

It is not clear how comparable measures of fatigue such as subjective ratings of ocular discomfort are with inferences drawn from performance rates. It would seem safe to conclude that users do not find reading from VDUs intrinsically fatiguing but that performance levels may be more difficult to sustain over time when reading from average quality screens. As screen standards increase over time this problem should be minimised if not eliminated.

2.4.4 Comprehension

Perhaps more important than the questions of speed and accuracy of reading is the effect of presentation medium on comprehension. Should any causal relationship ever be identified between reading from VDU and reduced comprehension, the impact of this technology would be severely limited. The issue of comprehension has not been as fully researched as one might expect, perhaps in no small way due to the difficulty of devising a suitable means of quantification i.e., how does one measure reader comprehension?

Post-task questions about content of the reading material are perhaps the simplest method of assessment, although care must be taken to ensure that the questions do not simply demand recall skills. Muter *et al.* (1982) required subjects to answer 25 multiple-choice questions after two 1 hour reading sessions. Due to variations in the amount of material read by all subjects, analysis was reduced to responses to the first eight questions of each set. No effect on comprehension was found either for condition or question set. Kak (1981) presented subjects with a version of the Nelson-Denny standardised reading test on paper and VDU. Comprehension questions were answered

by hand. No significant effect for presentation medium was observed. A similar result was found by Cushman (1986) in his comparison of paper, microfiche and VDUs. Interestingly however, he noted a negative correlation between reading speed and comprehension, i.e., comprehension tended to be higher for slower readers.

Belmore (1985) asked subjects to read short passages from screen and paper and measured reading time and comprehension using post-task questions. An initial examination of the results appeared to show a considerable disadvantage, in terms of both comprehension and speed, for screen presented text. However, further analysis showed that the effect was only found when subjects experienced the screen condition first. Belmore suggested that the performance decrement was due to the subjects' lack of familiarity with computers and reading from screens - a factor commonly found in this type of study. Very few of the studies reported here attempted to use a sample of regular computer users.

The Egan *et al.* study (1989) described earlier required subjects to write essay type answers to open book questions using paper or hypertext versions of a statistics book. Experts rated the essays and it was observed that users of the hypertext version scored significantly higher marks than users of the paper book. Thus, the authors conclude, the potential of restructuring the text with current technology can significantly improve comprehension for certain tasks.

It seems that comprehension of material is not negatively affected by electronic presentation and under some circumstances may even be improved. However, a strong qualification of this interpretation of the experimental findings is that suitable comprehension measures for reading material are difficult to devise. The expert rating used by Egan *et al.* is ecologically valid in that it conforms to the type of assessment usually employed in schools and colleges but the sensitivity of post-task question and answer sessions to subtle cognitive differences caused by presentation medium is debatable. Without evidence to the contrary though, it would seem as if reading from VDUs does not negatively affect comprehension.

2.4.5 Preference

Part of the folklore of human factors research is that naïve users tend to dislike using computers and much research aims at encouraging user acceptance of systems through more usable interface design. Given that much of the evidence cited here is based on studies of relatively novice users it is possible that the results are contaminated by

subjects' negative predispositions towards reading from screen. On the basis of a study of 800 VDU operators' comparisons of the relative qualities of paper and screen based text, Cakir *et al.* (1980) report that high quality typewritten hardcopy is generally judged to be superior. Preference ratings were also recorded in the Muter *et al.* (1982) study and despite the rather artificial screen reading situation tested, users only expressed a mild preference for reading from a book. They expressed the main advantage of book reading to be the ability to turn back pages and re-read previously read material, mistakenly assuming that the screen condition prevented this.

Starr (1984) concluded that relative subjective evaluations of VDUs and paper are highly dependent on the quality of the paper document (i.e., its typographic reproduction), though one may add that the quality of the VDU display probably has something to do with it too. Egan *et al.* (1989) found a preference for hypertext over paper amongst subjects in their study of a statistics text where the electronic copy was displayed on a very high quality screen. A similar finding was reported by Simpson (1989) who compared a small hypertext database with a collection of identical paper documents. However, lest it be concluded that the emergence of high quality screen is sufficient to overcome preference deficits for electronic text, Dillon (1988) found that subjects who were unable to satisfactorily manipulate electronic text rated a system with a very high quality screen as poorer than paper.

What seems to have been overlooked as far as formal investigation is concerned is the natural flexibility of books and paper over VDUs, e.g., books are portable, cheap, apparently "natural" in our culture, personal and easy to use. The extent to which such "common-sense" variables influence user performance and preferences is not yet well-understood but it is likely that they will play a significant role in the acceptance of any emerging technology.

2.4.6 Summary

Empirical investigations of the area have suggested five possible outcome differences between reading from screens and paper. As a result of the variety of methodologies, procedures and stimulus materials employed in these studies, definitive conclusions cannot be drawn. It seems certain that reading speeds are reduced on typical VDUs and accuracy may be lessened for cognitively demanding tasks. Fears of increased visual fatigue and reduced levels of comprehension as a result of reading from VDUs would however seem unfounded. With respect to reader preference, top quality hardcopy seems to be preferred to screen displays, which is not altogether surprising.

2.5 Process measures

Without doubt, the main obstacle to obtaining accurate process data is devising a suitable, non-intrusive observation method. While techniques for measuring eye-movements during reading now exist (and have done for several years), it is not at all clear from eye-movement records what the reader was thinking or trying to do at any time. Furthermore, use of such equipment is rarely non-intrusive, often requiring the reader to remain immobile through the use of head restraints, bite bars etc. or read the text one line at a time from a computer display —hardly equatable to normal reading conditions!

Less intrusive methods such as the use of light pens in darkened environments to highlight the portion of the text currently viewed (Whalley and Fleming 1975) or modified reading stands with semisilvered glass which reflect the reader's eye movements in terms of current text position to a video camera (Pugh 1979) are examples of the lengths researchers have gone to in order to record the reading process. However, none of these are ideal as they alter the reading environment, sometimes drastically, and only their staunchest advocate would describe them as non-intrusive.

Verbal protocols of people interacting with texts require no elaborate equipment and can be elicited wherever a subject normally reads. In this way they are cheap, relatively naturalistic and physically non-intrusive. However, the techniques have been criticised for interfering with the normal cognitive processing involved in task performance and requiring the presence of an experimenter to sustain and record the verbal protocol (Nisbett and Wilson 1977).

Although a perfect method does not yet exist it is important to understand the relative merits of those that are available. Eye-movement records have significantly aided theoretical developments in modeling reading (see e.g., Just and Carpenter 1980) while use of the light-pen-type techniques have demonstrated their worth in identifying the effects of various typographic cues on reading behaviour (see e.g. Waller 1984). Verbal protocols have been effectively used by researchers to gain information on reading strategies (see e.g. Olshavsky 1977).

Nevertheless, such techniques have rarely been employed with the intention of assessing the process differences between reading from paper and from screen. Where paper and hypertext are directly compared, although process measures may be taken

with the computer and/or video cameras, the final comparison often rests on outcome measures (e.g. McKnight *et al.*, 1990a).

Despite this, it is widely accepted that the reading process with screens is different from that with paper regardless of any outcome differences. The following sections outline three of the most commonly cited process differences between the media. In contrast to the outcome differences it will be noted that, for the reasons outlined above, these differences are less clearly demonstrated empirically.

2.5.1 Eye movements

Mills and Weldon (1985) argue that measures of eye movements reflect difficulty, discriminability and comprehensibility of text and can therefore be used as a method of assessing the cognitive effort involved in reading text from paper or screen. Indeed Tinker (1958) reports on how certain text characteristics affect eye movements and Kolers *et al.* (1981) employed measures of eye movement to investigate the effect of text density on ocular work and reading efficiency. Obviously if reading from screen is different than paper then noticeable effects in eye movement patterns might be found indicating possible causes and means of improvement.

Eye movements during reading are characterised by a series of jumps and fixations. The latter are of approximately 250 msec. duration and it is during these that word perception occurs. The 'visual reading field' is the term used to describe that portion of foveal and parafoveal vision from which visual information can be extracted during a fixation and in the context of reading this can be expressed in terms of the number of characters available during a fixation. The visual reading field is subject to interference from text on adjacent lines the effect of which seems to be a reduction in the number of characters available in any given fixation and hence a reduction in reading speed.

Gould *et al.* (1987a) report an investigation of eye movement patterns when reading from either medium. Using a photoelectric eye movement monitoring system, subjects were required to read two 10-page articles, one on paper, the other on screen. Eye movements typically consisted of a series of fixations on a line, with re-fixations and skipped lines being rare. Movement patterns were classified into four types: fixations, undershoots, regressions and re-fixations. Analysis revealed that when reading from VDU subjects made significantly more (15%) forward fixations per line. However this 15% difference translated into only 1 fixation per line. Generally, eye movement patterns were similar and no difference in duration was observed. Gould *et al.*

explained the 15% fixation difference in terms of image quality variables. Interestingly they report that there was no evidence that subjects lost their place,"turned-off" or re-fixated more when reading from VDUs.

It seems therefore that gross differences in eye movements do not occur between screen and paper reading. However, given the known effect of typographic cueing on eye movements with paper and the oft-stated non-transferability of paper design guidelines to screens, it is possible that hypertext formats might influence the reading process at this level in a manner worthy of investigation.

2.5.2 Manipulation

Perhaps the most obvious difference between reading from paper and from screens is the ease with which paper can be manipulated and the corresponding difficulty of so doing with electronic text. Manipulating paper is achieved by manual dexterity, using fingers to turn pages, keeping one finger in a section as a location aid, or flicking through tens of pages while browsing the contents of a document, activities difficult or impossible to support electronically (Kerr 1986).

Such skills are acquired early in a reader's life and the standard physical format of most documents means these skills are transferable between all document types. With electronic text this does not hold. Lack of standards means that there is a bewildering range of interfaces to computer systems and mastery of manipulation in one application is no guarantee of an ability to use another. Progressing through the electronic document might involve using a mouse and scroll bar in one application and function keys in another; one might require menu selection and "page" numbers while another supports touch-sensitive "buttons". With hypertext, manipulation of large electronic texts can be rapid and simple while other systems might take several seconds to refresh the screen after the execution of a "next page" command.

Such differences will almost certainly affect reading. Waller (1986) suggests that as readers need to articulate their needs in manipulating electronic texts (i.e., formulate an input to the computer to move the text rather than directly and automatically performing the action themselves) a distraction of cognitive resources required for comprehension could occur. Richardson *et al.* (1988) report that subjects find text manipulation on screen awkward compared to paper, stating that the replacement of direct manual interaction with an input device deprived users in their study of much feedback and control.

It is obvious that manipulation differences exist and that electronic text is usually seen as the less manipulable medium. Current hypertext applications however support rapid movement between various sections of text which suggests that innovative manipulations might emerge that, once familiar with them, convey advantages to the reader of electronic texts. This is an area for further work.

2.5.3 Navigation

There is a striking consensus among many researchers in the field that navigation is the single greatest difficulty for users of electronic text. This is particularly (but not uniquely) the case with hypertext where frequent reference is made to "getting lost in hyperspace" (e.g. Conklin 1987, McAleese 1989) which is described, in the oft-quoted line of Elm and Woods (1985), as:

> "the user not having a clear conception of the relationships within the system or knowing his present location in the system relative to the display structure and finding it difficult to decide where to look next within the system" (p. 927).

In other words, users do not know how the information is organised, how to find the information they seek or even if that information is available. With paper documents there tends to be at least some standards or conventions in terms of organisation. With books for example, contents pages are usually at the front, indices at the back and both offer some information on where items are located in the body of the text. Concepts of relative position in the text such as 'before' and 'after' have tangible physical correlates. No such correlation holds with hypertext and cues aiding positioning of material are greatly diminished in standard (i.e., non-hypertext) electronic text.

There is some direct empirical evidence in the literature to support the view that navigation can be a problem.[4] Edwards and Hardman (1989) for example, describe a study which required subjects to search through a specially designed hypertext. In total, half the subjects reported feeling lost at some stage.[5] Such feelings were mainly due to "not knowing where to go next" or "not knowing where they were in relation to the overall structure of the document" rather than "knowing where to go but not knowing

---

[4] Interestingly there is significantly less evidence than there are claims about navigation difficulties. Whether this reflects a view that navigation is such an obvious problem that it is not worth demonstrating or an over-willing acceptance of it as a problem on the basis of limited evidence is open to argument.

[5] This proportion is deduced from the data reported.

how to get there" (descriptors provided by the authors). Unfortunately, without direct comparison of ratings from subjects reading a paper equivalent we cannot be sure such proportions are solely due to using hypertext. However it is unlikely that many readers of paper texts do not know where they are in relation to the rest of the text![6]

Indirect evidence comes from the numerous studies which have indicated that users have difficulties with a hypertext (Monk *et al.*, 1988, Gordon *et al.*, 1988). Hammond and Allinson (1989) speak for many when they say:

> "Experience with using hypertext systems has revealed a number of problems for users..... First, users get lost... Second, users may find it difficult to gain an overview of the material... Third, even if users know specific information is present they may have difficulty finding it" (p. 294).

There are a few dissenting voices.[7] Brown (1988) argues that:

> "although getting lost is often claimed to be a great problem, the evidence is largely circumstantial and conflicting. In some smallish applications it is not a major problem at all" (p. 2).

This quote is telling in several ways. The evidence for navigational difficulties *is* often circumstantial, as noted above. The applications in which Brown claims it is not a problem at all, are, to use his word, "smallish" and this raises a crucial issue with respect to electronic text. When we speak of documents being so small a reader cannot 'get lost' in them or so large that navigation aids are required to use them effectively, the implication is that information occupies 'space' through which readers 'travel' or 'move'. Hammond and Allinson (1987) talk of the "travel metaphor" as a way of moving through a hypertext. Canter *et al.* (1985) speak of "routes through" a database. Even the dissenters believe that the reader or user navigates through the document, the only disagreement being the extent to which getting lost is a regular and/or serious occurrence.

The weight of evidence, be it experiential, anecdotal or empirical suggests that navigation is an issue worthy of consideration. Popular belief and some empirical work

---

[6] This descriptor has a certain ambiguity that confuses the issue. Positionally it is easy for the reader to know where she is in terms of the front, back or middle of the book. The same relationship is much more complex when we consider "where am I?" in relation to the argument.

[7] At a recent international workshop on hypermedia a well-known figure in the area stated emphatically during his presentation that "there is no navigation problem"! Unfortunately he neither produced evidence to support this statement nor repeated it in his written paper.

suggests that electronic text is harder to navigate than paper.

2.5.4 Summary

The reading process is affected by the medium of presentation though it is extremely difficult to quantify and demonstrate such differences empirically. The major differences are in manipulation which seems more awkward with electronic texts and navigation which appears to be more difficult with electronic and particularly hypertexts. Eye movement patterns do not seem to be significantly altered by presentation medium.

2.6 Explaining the differences: a classification of issues

While the precise nature and extent of the differences between reading from either medium have not been completely defined, attempts to identify possible causes of any difference have been made. A significant literature exists on issues dealing with display characteristics such as line length and spacing. It is not the aim of this review to detail this literature fully except where it relates to possible causes for reading differences between paper and screen that need to be addressed in designing electronic texts. Experimental investigations which have controlled such variables have still found performance deficits on VDUs, thus suggesting that the root cause of observed differences lies elsewhere. For a comprehensive review of these issues see Mills and Weldon (1985).

Examining the last 15 years of human factors research in this area it is possible to distinguish three types of investigation. Initial work concentrated on what could be termed "basic ergonomics" such as screen angle, image polarity and so forth. This work continues to some extent today. As technology developed and user interfaces afforded more sophisticated interaction with electronic texts issues to do with manipulation, such as scrolling versus paging, came to the fore. This is still an area of concern for many researchers but has become secondary in interest now to the issues of information structuring and navigation that have resulted from the emergence of hypertext.

In a very real sense all these areas are inter-related. Hypertext, by necessity involves reading from screens and manipulating electronic text and therefore research at the basic ergonomic level has relevance to the information structuring work, if only as a reminder of current technological limitations. Furthermore, such distinctions map neatly to the

distinction between perceptual, motor and cognitive levels of discourse, all of which are involved to some extent in reading. The following sections cover each of these work areas in turn before a framework relating these together is presented.

2.7 Basic ergonomic issues

An electronic text is physically different from a paper one. Consequently, many researchers have examined these aspects of the medium in an attempt to explain the performance differences. An exhaustive programme of work conducted by Gould and his colleagues at IBM between 1982 and 1987 represents probably the most rigourous and determined research effort. They tried to isolate a single variable responsible for observed differences. The following sections review this work and related findings in the search for an explanation of the observed performance differences between reading from paper and reading from VDUs.

2.7.1 Orientation

One of the advantages of paper over VDUs is that it can be picked up and orientated to suit the reader. VDUs present the reader with text in a reasonably rigid vertical orientation, though thanks to ergonomic design principles some flexibility to alter orientation is available in many systems. Gould *et al.* (1987a) investigated the hypothesis that differences in orientation may account for differences in reading performance. Subjects were required to read three articles, one on VDU, one on paper-horizontal and the other on paper-vertical. Both paper conditions were read significantly faster than the VDU and there were no accuracy differences. While orientation has been shown to affect reading rate of printed material (Tinker, 1963) it does not appear to explain the observed reading differences in the comparisons reported here.

2.7.2 Visual angle

Gould *et al.* (1987a) hypothesised that due to the usually longer line lengths on VDUs the visual angle subtended by lines on each medium differs and that people have learned to compensate for the longer lines on VDUs by sitting further away from them when reading. In an initial crude experiment of reading differences Gould and colleagues visited the offices of 26 people who were reading either from VDU or paper and measured reading distance from both media with a metre stick. They found significantly greater reading distances for VDUs. Further work has confirmed that preferred viewing distance for screens is greater than that for paper (Jaschinski-Kruza 1990).

In a more controlled follow-up study Gould *et al.* (1987a) had 18 subjects read twelve different three-page articles for misspellings. Subjects read two articles at each of six visual angles: 6.7, 10.6, 16.0, 24.3, 36.4 and 53.4 degrees. Results showed that visual angle significantly affected speed and accuracy. However the effects were only noticeable for extreme angles, and between a range of 16.0 to 36.4 degrees, which covers typical VDU viewing, no effect for angle was found.

2.7.3 Aspect ratio

The term aspect ratio refers to the relationship of width to height. Typical paper sizes are higher than they are wider, while the opposite is true for typical VDU displays. Changing the aspect ratio of a visual field may affect eye movement patterns sufficiently to account for some of the performance differences. Gould *et al.* (1987a) had 18 subjects read three eight-page articles on VDU, paper and paper-rotated (aspect ratio altered to resemble screen presentation). The results however showed little effect for ratio.

2.7.4 Dynamics

Detailed work has been carried out on screen filling style and rates (e.g., Bevan, 1981; Kolers *et al.*, 1981; Schwartz *et al.*, 1983) and findings suggest that variables such as rate and direction of scrolled text do influence performance and subjective ratings. In order to understand the role of dynamic variables such as scrolling, "jittering" and screen filling in reading from VDUs, Gould *et al.* (1987a) had subjects read from paper, VDU and good quality photographs of the VDU material which eliminated any dynamics. Results provided little in the way of firm evidence to support the dynamics hypothesis. Subjects again read consistently faster from paper compared to both other presentation media, which did not differ significantly from each other. Creed *et al.* (1987) also compared paper, VDU and photos of the screen display on a proofreading task with 30 subjects. They found that performance was poorest on VDU but photographs did not differ significantly from either paper or VDU in terms of speed or accuracy, though examination of the raw data suggested a trend towards poorer performance on photos than paper. It seems unlikely therefore that much of the cause for differences between the two media can be attributed to the dynamic nature of the screen image.

2.7.5 Flicker

Characters are written on a VDU by an electron beam which scans the phosphor surface of the screen, causing stimulated sections to glow temporarily. The phosphor is characterised by its persistence, a high-persistence phosphor glowing for longer than a low-persistence phosphor. In order to generate a character that is apparently stable it is necessary to rescan the screen constantly with the requisite pattern of electrons. The frequency of scanning is referred to as the refresh rate since it is effectively refreshing the screen contents. Since the characters are in effect repeatedly fading and being regenerated it is possible that they appear to flicker rather than remain constant. The amount of perceived flicker will obviously depend on both the refresh rate and the phosphor's persistence; the more frequent the refresh rate and the longer the persistence, the less perceived flicker. However refresh rate and phosphor persistence alone are not sufficient to predict whether or not flicker will be perceived by a user. It is also necessary to consider the luminance of the screen. While a 30 Hz refresh rate is sufficient to eliminate flicker at low luminance levels, Bauer *et al.* (1983) suggested that a refresh rate of 93 Hz was necessary in order for 99% of subjects to perceive a positive presentation (dark characters on light background) display as flicker free.

If flicker was responsible for the large differences between reading from paper and VDU it would be expected that studies such as Creed *et al.*'s (1987) which employed photographs of screen displays would have demonstrated a significant difference between reading from photos and VDUs. However the extent to which flicker may have been an important variable in many studies is unknown as details of screen refresh rates are often not included in publications. Gould *et al.* (1987a) admit that the photographs used in their study were of professional quality but appeared less clear than the actual screen display. It is likely that using photos to control flicker may not be a suitable method and flicker may play some part in explaining the differences between the two media.

2.7.6 Image polarity

As briefly mentioned earlier, a display in which dark characters appear on a light background (e.g., black on white) is referred to as positive image polarity or negative contrast. This will be referred to here as positive presentation. A display on which light characters appear on a dark background (e.g., white on black) is referred to as negative image polarity or positive contrast. This will be referred to here as negative

presentation. The traditional computer display involves negative presentation, typically white on black though light green on dark green is also common.

Since 1980 there has been a succession of publications concerned with the relative merits of negative and positive presentation. Several studies suggest that, tradition notwithstanding, positive presentation may be preferable to negative. For example Radl (1980) reported increased performance on a data input task for dark characters and Bauer and Cavonius (1980) reported a superiority of dark characters on various measures of typing performance and operator preference.

With regards to reading from screens Cushman (1986) reported that reading speed and comprehension on screens was unaffected by polarity, though there was a non-significant tendency for faster reading of positive presentation. Gould *et al.* (1987b) specifically investigated the polarity issue. Fifteen subjects read 5 different 1000 word articles, 2 negatively presented, 2 positively presented and one on paper (standard positive presentation). Further experimental control was introduced by fixing the display contrast for one article of each polarity at a contrast ratio of 10:1 and allowing the subject to adjust the other article to their own liking. This avoided the possibility that contrast ratios may have been set which favoured one display polarity. Results showed no significant effect for polarity or contrast settings, though 12 of the 15 subjects did read faster from positively than negatively presented screens, leading the investigators to conclude that display polarity probably accounted for some of the observed differences in reading from screens and paper.

In a general discussion of display polarity Gould *et al.* (1987b) state that:

> "to the extent that polarity makes a difference it favours faster reading from dark characters on a light background" (p.514).

Furthermore they cite Tinker (1963) who reported that polarity interacted with type size and font when reading from paper. The findings of Bauer *et al.* (1983) with respect to flicker certainly indicate how perceived flicker can be related to polarity. Therefore the contribution of display polarity in reading from screens is probably important through its interactive effects with other display variables.

2.7.7 Display characteristics

Issues related to fonts such as character size, line spacing and character spacing have

been subjected to detailed research. However the relationship of much of the findings to reading continuous text from screens is not clear.

Character size on VDUs is closely related to the dimension of the dot matrix from which the characters are formed. Traditionally 5x7 matrices have been used but they offer little opportunity for representing lower-case ascenders and descenders, and consequently produce poor legibility. The dramatic increase in computer processing power now means that there is little cost in employing larger matrices and Cakir *et al.* (1980) recommend a minimum of 7x9. Pastoor *et al.* (1983) studied the relative suitability of four different dot-matrix sizes and found reading speed varied considerably. On the basis of these results the authors recommended a 9x13 character size matrix. However their study was concerned with television screens and their tasks included isolated word reading and column searching. In short, the optimum character size for reading from screens appears to be contingent on the task performed although there is likely to be a minimum size imposed by screen resolution.

Considerable experimental evidence exists to favour proportionally rather than non-proportionally spaced characters (e.g., Beldie *et al.*, 1983). Once more though, the findings must be viewed cautiously. In the Beldie *et al.* study for example, the experimental tasks did not include reading continuous text. Muter *et al.* (1982) compared reading speeds for text displayed with proportional or non-proportional spacing and found no effect. In an experiment intended to appreciate the possible effect of such font characteristics on the performance differences between paper and screen reading Gould*et al.* (1987b) found no evidence to support the case for proportionally spaced text.

Kolers *et al.*(1981) studied interline spacing and found that with single spacing significantly more fixations were required per line, fewer lines were read and the total reading time increased. However the differences were small and were regarded as not having any significance. On the other hand Kruk and Muter (1984) found that single spacing produced 10.9% slower reading than double spacing. Once more the results appear inconclusive.

Obviously much work needs to be done before a full understanding of the relative advantages and disadvantages of particular formats and types of display is achieved. In a discussion of the role of display fonts in explaining any of the observed differences between screen and paper reading Gould *et al.* (1987b) state that on the basis of their investigations there is "strong evidence that font (within reason) has little effect on

reading rate from paper" (p.515). They add that it is almost impossible however to discuss fonts without recourse to the physical variables of the computer screen itself e.g., screen resolution and beam size, once more highlighting the potential cumulative effect of several interacting factors on reading from screens

2.7.8 Anti-aliasing

Most computer displays are raster displays typically containing dot matrix characters and lines which give the appearance of "staircasing" i.e. edges of characters may appear jagged. This is caused by undersampling the signal that would be required to produce sharp, continuous characters. The process of anti-aliasing has the effect of perceptually eliminating this phenomenon on raster displays. A technique for anti-aliasing developed by IBM accomplishes this by adding variations in grey level to each character.

The advantage of anti-aliasing lies in the fact that it improves the quality of the image on screen and facilitates the use of fonts more typical of those found on printed paper. To date the only reported investigation of the effects of this technique on reading from screens is that of Gould *et al.* (1987b). They had 15 subjects read three different 1000 word articles, one on paper, one on VDU with anti-aliased characters and one on VDU without anti-aliased characters. Results indicated that reading from anti-aliased characters did not differ significantly from either paper or non anti-aliased characters though the latter two differed significantly from each other. Although the trend was present the results were not conclusive and no certain evidence for the effect of anti-aliasing was provided. However the authors report that 14 of the 15 subjects preferred the anti-aliased characters, describing them as clearer and easier to read.

2.7.9 User characteristics

It has been noted that many of the studies reported in this review employed relatively naïve users as subjects. The fact that different types of users interact with computer systems in different ways has long been recognised and it is possible that the differences in reading that have been observed in these studies result from particular characteristics of the user group involved.

Most obviously, it might be assumed that increased experience in reading from computers would reduce the performance deficits. A direct comparison of experienced and inexperienced users was incorporated into a study on proofreading from VDUs by

Gould *et al.* (1987a). Experienced users were described as regular, daily users who had worked with VDUs for a year or more. Inexperienced users had no experience of reading from computers. No significant differences were found between these groups, both reading slower from screen.

No reported differences for age or sex can be found in the literature. Therefore it seems reasonable to conclude that basic characteristics of the user are not responsible for the differences in reading from these presentation media.

2.7.10 The interaction of display variables: the work of Gould

Despite many of the findings reported thus far, it appears that reading from screens can at least be as fast and as accurate as reading from paper. Gould *et al.* (1987b) have empirically demonstrated that under the right conditions some of the differences between the two presentation media disappear. In a study employing sixteen subjects, an attempt was made to produce a screen image that closely resembled the paper image i.e., similar font, size, colouring, polarity and layout were used. Univers-65 font was positively presented on a monochrome IBM 5080 display with an addressability of 1024x1024. No significant differences were observed between paper and screen reading. This study was replicated with twelve further subjects using a 5080 display with an improved refresh rate (60Hz). Again no significant differences were observed though several subjects still reported some perception of flicker.

On balance it appears that any explanation of these results must be based on the interactive effects of several of the variables outlined in the previous sections. After a series of experimental manipulations aimed at identifying those variables responsible for the improved performance Gould *et al.* (1987b) suggested that the performance deficit was the product of an interaction between a number of individually non-significant effects. Specifically, they identified display polarity (dark characters on a light, whitish background), improved display resolution, and anti-aliasing as major contributions to the elimination of the paper/screen reading rate difference.

Gould *et al.* (1987b) conclude that the explanation of many of the reported differences between the media is basically visual rather than cognitive and lies in the fact that reading requires discrimination of characters and words from a background. The better the image quality is, the more reading from screen resembles reading from paper and hence the performance differences disappear. This seems an intuitively sensible conclusion to draw. It reduces to the level of simplistic any claims that one or other

variable such as critical flicker frequency, font or polarity are responsible for any differences.

The Gould *et al.* (1987b) findings are of tremendous importance. They would suggest that the results of the many studies reported earlier can be explained in terms of the quality of screen image presented to the subjects. Muter *et al.* (1982), for example, employed television screens with negative presentation in their investigation. Wilkinson and Robinshaw (1987) also used negative presentation and a screen described by themselves as "of average quality". In fact none of the studies reporting performance deficits that are cited in this review can claim to have presented screen images of the quality employed in the Gould *et al.* (1987b) studies.

2.7.11 Conclusion

Although reading from computer screens may be slower and occasionally less accurate than reading from paper, no one variable is likely to be responsible for this difference. It is almost certain that neither inherent problems with the technology nor the reader are causal factors. Invariably it is the quality of the image presented to the reader which is crucial. Tinker (1963) reports dramatic interaction effects of image quality variables on paper and according to Gould *et al.* (1987b) it is likely that these occur on screen too. Positive presentation combined with a high screen resolution with high refresh rate to avoid flicker can produce good images and with the addition of anti-aliased characters it becomes possible to provide a screen display that resembles the print image and thereby facilitates reading. It must be remembered however that typical computer displays present images that are still of poorer quality than those used by Gould and his associates to overcome the performance deficit. Until screen standards are raised sufficiently these differences are likely to remain.

A major shortcoming of the studies by Gould *et al.* is that they only address limited outcome variables: speed and accuracy. Obviously speed is not always a relevant criterion in assessing the output of a reading task. Furthermore, the accuracy measures taken in these studies have been criticised as too limited and further work needs to be carried out to appreciate the extent to which the explanation offered by Gould is sufficient. It follows that other observed outcome differences such as fatigue and reader preference should also be subjected to investigation in order to understand how far the image quality hypothesis can be pushed as an explanation for reading differences between the two media.

A shortcoming of all the work cited in this section is the task employed. Invariably it was proofreading or some similar task which hardly constitutes normal reading for most people. Thus the ecological validity of these studies is low. Beyond this, the actual texts employed were all relatively short (Gould's for example averaged only 1100 words but many other researchers used even shorter texts). As a result, it is difficult to generalise these conclusions beyond the specifics of task and texts employed to the wider class of activities termed "reading". Creed *et al.* (1987) defend the use of proofreading on the grounds of its amenability to manipulation and control. While this desire for experimental rigour is laudable one cannot but feel that the major issues involved in using screens for real-world reading scenarios are not addressed by such work. With this in mind, the following section considers the literature on research concerned with the manipulation facilities.

2.8 Manipulation facilities

It is clear that the search for the specific ergonomic variables responsible for differences between the media has been insightful. However, few readers of electronic texts would be satisfied with the statement that the differences between the media are visual rather than cognitive. This might explain absolute speed and accuracy differences on limited tasks but hardly accounts for the range of process differences that are found as described earlier.

Once the document becomes too large to display on a single screen other factors than image quality immediately come into play. Several researchers have pinned their hopes on improved manipulation facilities with electronic texts removing many of the differences between the media, or at least compensating for lower image quality by offering faster access to reading material. In this section, research into variables affecting such issues is reviewed.

2.8.1 Scrolling versus paging

The manner in which a reader moves through a document is distinctly different in either medium and even within the electronic medium, various techniques are employed for displaying sections of the text. Scrolling and paging are two of the most common.

There is evidence to suggest that readers establish a visual memory for the location of items within a printed text based on their spatial location both on the page and within

the document (Rothkopf, 1971; Lovelace and Southall, 1983). This memory is supported by the fixed relationship between an item and its position on a given page. A scrolling facility is therefore liable to weaken these relationships and offers the reader only the relative positional cues that an item has with its immediate neighbours. However, on the basis of a literature review, Mills and Weldon (1985) report that there is no real difference between scrolling and paging though Schwartz *et al.* (1983) found that novices tend to prefer paging (probably based on its close adherence to the book metaphor) and Dillon *et al.* (1990a) report that a scrolling mechanism was the most frequently cited improvement suggested by subjects assessing their reading interface.

Scrolling has also been investigated in conjunction with direction (vertical or horizontal — Sekey and Tietz, 1982), rate (self-paced or machine-paced — Kolers *et al.*, 1981) and display size (Duchnicky and Kolers, 1983). With reference to direction and rate, all seem to agree that ideally, lengthy texts should be presented vertically and at the reader's choice of rate. Even so, Kolers *et al.* (1981) report that forcing readers to increase their rates by 10-20% does not lead to loss of comprehension and actually appears to increase efficiency of eye-movements as measured by rate and length of fixation.

It seems therefore that scrolling is a popular form of text manipulation with more experienced users probably due to its speed even if there are theoretical grounds for doubting its superiority over paging. There is no evidence that either facility significantly affects reading performance compared to paper.

2.8.2 Display size

Display size is a much discussed but infrequently studied aspect of human-computer interaction in general and reading electronic text in particular. Popular wisdom suggests that "bigger is better" but empirical support for this edict is sparse. Duchnicky and Kolers (1983) investigated the effect of display size on reading constantly scrolling text and reported that there is little to be gained by increasing display size to more than 4 lines either in terms of reading speed or comprehension. Elkerton and Williges (1984) investigated 1, 7, 13, and 19-line displays and reported that there were few speed or accuracy advantages between the displays of 7 or more lines. Similarly, Neal and Darnell (1984) report that there is little advantage in full page over partial page displays for text-editing tasks (which is only partly relevant to the concept of reading).

These results seem to suggest that there is some critical point in display size, probably around 5 lines, above which improvements are slight. Intuitively this seems implausible. Few readers of paper texts would accept presentations of this format. Our experiences with paper suggest that text should be displayed in larger units than this. Furthermore, loss of context is all too likely to occur with lengthy texts and the ability to browse and skim backward and forward is much easier with 30 or so lines of text than with 5 line displays. Of the experiments cited, only the Duchnicky and Kolers study was concerned with reading for comprehension and their passages were never longer than 300 words. Thus their findings on window size seem to bear little relevance to reading of lengthy texts.

Deliberately examining this, Richardson *et al.* (1989) had subjects perform 10 information location tasks using an electronic book with a display size of 20 or 40 lines. Though they observed no performance differences between conditions they did report a significant preference effect favouring the larger display. Similarly Dillon *et al.* (1990a) investigated screen sizes of 20 and 60 lines for reading an electronic version of an academic article. Interestingly they found a manipulation effect for screen size that could not be explained by the fact that to read a complete text on a small screen necessitates more manipulations than seeing it on a large one. They reported that when such simple manipulations are discounted and attention is paid only to changes in direction or jumps of 2 or more "pages", readers using the small screen still manipulated the text more. They proposed that the likeliest explanation was that readers like to re-read large parts of texts or jump about when using articles and that the smaller screen condition required more manipulations to observe the same amount of text as the bigger screen. As in the Richardson *et al.* study, the authors report a preference effect favouring the larger display.

As with many variables, the task being performed is likely to be a deciding factor. Small screens pose problems for readers wishing to browse through lengthy texts but are likely to be more acceptable for tasks requiring a straight perusal of short material such as a letter or memo. Significantly, many applications now allow the user to change window size within the constraints of the overall screen size which may accommodate some preference differences but doesn't resolve issues to do with ultimate screen size.

It is likely that many of the effects of screen size are too subtle to be assessed by gross outcome measures such as speed and accuracy. Larger screens might suit better spatial memory formation or browsing, variables that are not usually measured by investigators. As concluded in the basic ergonomic research, it is likely that the

interaction of size with other manipulation variables is important.

2.8.3 Text splitting across screens

A related issue to display size and scrolling/paging is the splitting of paragraphs mid-sentence across successive screens. In this case, which is more likely to occur in small displays, the reader must manipulate the document in order to complete the sentence. This is not a major issue for paper texts such as books or journals because (unlike theses) the reader is usually presented with two pages at a time and access to previous pages is normally easy. On screen however, access rates are not so fast and the break between screens of text is more likely to be critical.

Research into reading has clearly demonstrated the complexity of the cognitive processing that occurs. The reader does not simply scan and recognise every letter in order to extract the meaning of words and then sentences. Comprehension requires inference and deduction, and the skilled reader achieves much of his smoothness by predicting probable word sequences (Chapman and Hoffman, 1977). The basic units of comprehension in reading that have been proposed are propositions (Kintsch, 1974), sentences (Just and Carpenter, 1980) and paragraphs (Mandler and Johnson, 1977). Splitting sentences across screens is likely to disrupt the process of comprehension by placing an extra burden on the limited capacity of working memory to hold the sense of the current conceptual unit while the screen is filled. Furthermore, the fact that between 10-20% of eye movements in reading are regressions to earlier fixated words and that significant eye movement pauses occur at sentence ends would suggest that sentence splitting is also likely to disrupt the reading process and thereby hinder comprehension.

In the Dillon *et al.* (1990a) study cited earlier, the role of text splitting on performance was also examined. They found that splitting text across screens caused readers to return to the previous page to re-read text significantly more often than when text was not split. Though this appeared to have no effect on subsequent comprehension of the material being read, they concluded that it was remarked upon by the subjects sufficiently often to suggest that it would be a nuisance to regular users. In this study however the subjects were reading from a paging rather than scrolling interface where the effect of text splitting was more likely to cause problems due to screen-fill delays. With scrolling interfaces text is always going to split across screen boundaries but there is rarely a perceptible delay in image presentation to disrupt the reader. In sum therefore, it seems as though text splitting should be avoided for paging interfaces.

2.8.4 Window format

It has become increasingly common to present information on computer screen via windows i.e., sections of screen devoted to specific groupings of material. Current technology supports the provision of independent processes within windows or the linking of inputs in one window with the subsequent display in another, the so-called "co-ordinated windows" approach (Shneiderman, 1987).

Such techniques have implications for the presentation of text on screen as they provide alternatives to the straightforward listing of material in "scroll" form or as a set of "pages". For example, while one window might present a list of contents in an electronic text, another might display whole sections of it according to the selection made. In this way, not only is speed of manipulation increased but the reader can be provided with an overview of the document's structure to aid orientation while reading an opened section.

The use of such techniques is now commonplace in hypertext applications. GUIDE for example, uses windows in one instance to present short notes or diagrams as elaborations or explanations of points raised in the currently viewed text, rather like sophisticated footnotes. The concept of hypertext as non-linear text is, in a very real sense, derived from such presentation facilities.

Tombaugh *et al.* (1987) investigated the value of windowing for readers of lengthy electronic texts. They had subjects read two texts on single or multi-window formats before performing 10 information location tasks. They found that novices initially performed better with a single-window format but subsequently observed that, once familiar with the manipulation facilities, the benefits of multi-windowing in terms of aiding spatial memory became apparent. They highlight the importance of readers acquiring familiarity with a system and the concept of the electronic book in order to accrue the benefits of such facilities.

Simpson (1989) compared performance with a similar multi-window display, a tiled display (in which the contents of each window were permanently visible) and a 'conventional' stack of windows (in which the windows remained in reverse order of opening). She reported that performance with the conventional window stack was poorest but that there was no significant difference between the tiled and multi-window displays. She concluded that for information location tasks, the ability to see a window's contents is not as important as being able to identify a permanent location for

a section of text.

Both of these studies highlight the impact of display format on readers' performance of a standard reading tasks: information location in a body of text. Spatial memory seems important and paper texts are good at supporting its use. Windowing, if deployed so as to retain order can be a useful means of overcoming this inherent weakness of electronic text. However, studies examining the problems of windowing very long texts where more than five or six stacked windows are required must be carried out before any firm conclusions about the benefits of this technique can be drawn.

2.8.5 Search facilities

Electronic text supports word or term searches at rapid speed and with total accuracy and this is clearly an advantage for users in many reading scenarios, e.g., checking references, seeking relevant sections etc. Indeed it is possible for such facilities to support tasks that would place unreasonable demands on users of paper texts, e.g., searching a large book for a non-indexed term or several volumes of journals for references to a concept.

Typical search facilities require the user to input a search string and choose several criteria for the search such as ignoring certain text forms (e.g., all upper case words) but sophisticated facilities on some database systems can support specification of a range of texts to search. The usual form for search specification is Boolean, i.e., users must input search criteria according to formal rules of Boolean logic employing constructs such as 'not', 'or' as well as 'and', which when used in combination support powerful and precise specifications. Unfortunately most end-users of computer systems are not trained in their use and while the terms may appear intuitive, they are often difficult to employ successfully.

In current electronic text facilities a simple word search is most common but users still seem to have difficulties. Richardson *et al.* (1988) reported that subjects in their experiment displayed a tendency to respond to unsuccessful searches by increasing the specificity of the search string rather than lessening it. The logic appeared to be that the computer required precision rather than approximation to search effectively. While it is likely that such behaviour is reduced with increased experience of computerised searching, a study by McKnight *et al.* (1990a) of information location within text found other problems. Here, when searching for the term "wormwood" in an article, two subjects input the search term "woodworm," displaying the intrusion of a common

sense term for an unusual word of similar sound and shape (a not uncommon error in reading under pressure due to the predictive nature of this act during sentence processing). When the system correctly returned a "Not Found" message, the users concluded that the question was an experimental trick, displaying their belief that the system was foolproof and that they had input the search criterion correctly.

Thus it seems as if search facilities are a powerful means of manipulating and locating information on screen and convey certain advantages impossible to provide in the paper medium. However, users may have difficulties with them in terms of formulating accurate search criteria. This is an area where research into the design of search facilities and increased exposure of users to electronic information can lead to improvements resulting in a positive advantage of electronic text over paper.

2.8.6 Input device

Over the last 15 years numerous input devices have been designed and presented as optimum for users e.g., trackerball, mouse, function keyboard, joystick, light pen etc. Since Card *et al.*'s (1978) claim that the speed of text selection via a mouse was constrained only by the limits of human information processing, this device has assumed the dominant position in the market.

It has since become clear that, depending on the task, other input devices can significantly outperform the mouse (Milner 1988). For example, when less than ten targets are displayed on screen and the cursor can be made to jump directly from one to the next, cursor keys are faster than a mouse (Shneiderman 1987). In the electronic text domain, Ewing *et al.* (1986) found this to be case with the HyperTIES application, though there is reason to doubt their findings as the mouse seems to have been used on less than optimum surface conditions.

Though "direct manipulation" (Shneiderman, 1984) might be a common description of an interface, it seems that its current manifestations leave much to be desired when it comes to manipulating text. Obviously practice and experience will play a considerable part here. Expertise with an input device affords the user a high level of control and breeds a sense of immediacy between selection and action.

It is important to realise that the whole issue of input device cannot be separated from other manipulation variables such as scrolling or paging. For example, a mouse that must be used in conjunction with a menu for paging text will lead to different

performance characteristics than one used with a scroll bar. For the moment however the mouse appears dominant and as the "point and click" concept becomes integrated with the "look and feel" of hypertext it will prove difficult to replace, even if convincing experimental evidence against its use, or an innovative credible alternative should emerge. [8]

2.8.7 Icon design

Recent advances in interface design have led to the emergence of more graphical than textual control mechanisms for interacting with a computer. The "I" in WIMP[9] (a *de facto* contemporary standard if not a legislated one) stands for icon and it is common for users to perform a range of intricate and heretofore complex operations at the click of a mouse button on an icon.

There are sound theoretical grounds for supporting iconic representation. Being language independent icons convey information by pictographic means and should thus support use by individuals unfamiliar with the terminology of operating systems and command languages. Further advantages of iconic representations are that they utilise little display space and render syntax errors on input obsolete (Gittins 1986).

On the negative side, icons can be confusing if their form provides no immediate clue to their action. Trashcans and folders might be intuitive but this is not always the case (the "home" icon on HyperCard is a picture of a little house and naïve users have failed to appreciate the intended reference [McKnight *et al.*, 1989]). Designing icons to convey less obvious actions than "goto" is not a simple task. Some designers even provide icons with textual descriptors to provide clues to their use which seems to defeat the purpose.

In manipulating documents electronically, icons have become popular in many hypertext applications. GUIDE for example uses such forms as boxes, arrows and circles when the cursor moves over an actionable area of the document, while HyperCard provides numerous "button" shapes that cause different document

---

[8] One has only to consider the dominance of the far from optimum QWERTY keyboard to understand how powerful convention is. This keyboard was actually designed to slow human operators down by increasing finger travel distances thereby lessening key jamming in early mechanical designs. Despite evidence that other layouts can provide faster typing speeds with less errors (Martin 1972) the QWERTY format retains its dominant position.

[9] WIMP stands for Windows, Icons, Menus and Pointing device (or Windows, Icons, Mouse and Pull-down menus) and is the term used to describe the "look and feel" of certain (mainly the Apple Macintosh) graphical user interfaces.

manipulations to occur. Used in conjunction with a mouse such facilities can support rapid, easy manipulations of the text and allow the user to access the document through numerous routes – instantiating the notion of non-linearity in hypertext.

Icons are also used to represent a document in situations where the user might be selecting one of several texts. While it is easy enough to convey an image of book or other text type iconically few systems attempt to provide the range of cues available with paper such as size, age, level of usage and so forth.

Stammers *et al.* (1989) reported that icons are most useful when they represent concrete rather than abstract actions which while intuitively sensible, suggests ultimate limitations on their use as many computer functions are highly abstract in nature. Brems and Whitten (1987) found that icons were more appropriate for experienced than novice users which is ironic given the stated benefits of icons. Shulman *et al.* (1985) found no benefits from icons over names as command designators in text editing tasks which suggests that they many not be the panacea they are often claimed to be, and that more specific studies of their use in electronic text is called for.

Generalising such findings to the electronic text domain is difficult at present. A reasonable conclusion seems to be that icons have a role, particularly for simple or repetitive actions such as "go there" or "look at this in more detail" but are less applicable for conveying information of abstract actions. For manipulation purposes the basic range of actions is always likely to be limited therefore it is conceivable that standard designs for such actions might appear soon. More elaborate actions should utilise other command forms.

2.8.8 Conclusion

Manipulating electronic text is considered to be more difficult than manipulating paper. Research suggests that factors such as non-splitting of text, rapid response and increased display size can improve matters and that facilities such as searching and multi-windowing might even offer benefits to electronic text over paper.

As with the basic ergonomic issues reviewed earlier the interaction of several of these variables is likely to be crucial. Small displays limit windowing facilities and may increase text-splitting causing manipulation differences with paper that might not emerge with large, multi-windowed displays. Furthermore, as Tombaugh *et al.* (1987) pointed out, familiarity with the facilities is vital. It is not always clear from the

literature how this variable has been controlled in many studies.

The range of tasks used for such investigations is much wider and therefore more appropriate than those used in the basic ergonomic work reviewed. Some studies have been particularly valid both ecologically and as a test of manipulation facilities, therefore it is difficult to criticise this work on these grounds.

As an explanation of the differences between the media, manipulation must be incomplete however. Even if combined with good image quality, optimum manipulation facilities are unlikely to remove all the problems associated with electronic text. This is becoming obvious from much of the recent work on hypertext that is concerned with structuring information. Accordingly, research at this level is now considered.

2.9 Information structure

Once a document becomes larger than several screens it starts posing difficulties for readers in terms of locating material and remembering where they saw something. Research into basic ergonomic issues and manipulation facilities provide no answers here. Added to this, the emergence of hypertext and its scope for supporting alternative information structures to paper has led some researchers to consider these factors to be the most important ones in explaining and overcoming differences between the media.

The following sections describe some of the work and ideas proposed at this level of analysis. This is the level of least specificity in the conclusions that can be drawn and also the widest ranging in issues considered. Consequently much of the work here is discursive rather than empirical in nature.

2.9.1 Readers' models of paper and electronic documents

Exposure to the variety of texts in everyday life leads readers to acquire mental models or schemata for documents they are familiar with. Schemata are hypothetical knowledge structures which mentally represent the general attributes of a concept and are considered by psychologists to provide humans with powerful organising principles for information (Bartlett ,1932; Cohen, 1988). Thus, when we pick up a book we immediately have expectations about the likely contents. Inside the front cover we expect such details as where and when it was published, perhaps a dedication and then

a Contents page. We know, for example, that contents listings describe the layout of the book in terms of chapters, proceeding from the front to the back. Chapters are organised around themes and an index at the back of the book, organised alphabetically, provides more specific information on where information is located in the body of the text. Experienced readers know all this before even opening the text. It would strike us as odd if such structures were absent or their positions within the text were altered.

According to van Dijk and Kintsch (1983), such models or schemata, which they term 'superstructures', facilitate comprehension of material by allowing readers to predict the likely ordering and grouping of constituent elements of a body of text. To quote van Dijk (1980):

> "a superstructure is the schematic form that organises the global meaning of a text. We assume that such a superstructure consists of functional categories.. (and).. rules that specify which category may follow or combine with what other categories." (p.108).

But apart from categories and functional rules, van Dijk adds that a superstructure must be socioculturally accepted, learned, used and commented upon by most adult language users of a speech community. Research by van Dijk and Kintsch (1983) and Kintsch and Yarborough (1982) has shown how such structures influence comprehension of texts.

In this format the schema/superstructure constitutes a set of expectancies about their usual contents and how they are grouped and positioned relative to each other. In advance of actually reading the text readers cannot have much insight into anything more specific than this, but the generality of organisation within the multitude of texts read in everyday life affords stability and orientation in what could otherwise be a complex informational environment.

The concept of a schema for an electronic information space is less clear-cut than those for paper documents. Electronic documents have a far shorter history than paper and the level of awareness of technology among the general public is relatively primitive compared to that of paper. Exposure to information technology will almost certainly improve this state of affairs but even among the contemporary computer literate it is unlikely that the type of generic schematic structures that exist for paper documents

have electronic equivalents of sufficient generality.[10]

Obviously computing technology's short history is one of the reasons but it is also the case that the medium's underlying structures do not have equivalent transparency. Thus using electronic information is often likely to involve the employment of schemata for systems in general (i.e., how to operate them) in a way that is not essential for paper-based information.

The qualitative differences between the schemata for paper and electronic documents can easily be appreciated by considering what you can tell about either at first glance. The information available to paper text users was outlined above. When we open a hypertext or other electronic document however we do not have the same amount of information available to us. We are likely to be faced with a welcoming screen which might give us a rough idea of the contents (i.e., subject matter) and information about the authors/developers of the document but little else. Such displays are usually two-dimensional, give no indication of size, quality of contents, age (unless explicitly stated) or how frequently the text has been used (i.e., there is no dust or signs of wear and tear on it such as grubby finger-marks or underlines and scribbled comments).

Performing the electronic equivalent of opening up the text or turning the page offers no assurance that expectations will be met. Many hypertext documents offer unique structures (intentionally or otherwise) and their overall sizes are often impossible to assess in a meaningful manner (these points are dealt with in more detail in Dillon *et al.* 1990b). At their current stage of development it is likely that users/readers familiar with hypertext will have a schema that includes such attributes as linked nodes of information, non-serial structures, and perhaps, potential navigational difficulties! The manipulation facilities and access mechanisms available in hypertext will probably occupy a more prominent role in their schemata for hypertext documents than they will for readers' schemata of paper texts. As yet, empirical evidence for such schemata is lacking.

The fact that hypertext offers authors the chance to create numerous structures out of the same information is a further source of difficulty for users or readers. Since schemata are generic abstractions representing typicality in entities or events, the increased variance of hypertext implies that any similarities that are perceived must be at a higher level or must be more numerous than the schemata that exist for paper texts.

---

10 It is worth noting that in part, this might be because the electronic document is usually only a stage in the production of a paper one. Few pure electronic texts exist thus any unique forms have yet to emerge.

It seems therefore that users' schemata of electronic texts are likely to be "informationally leaner" than those for paper documents. This is attributable to the recent emergence of electronic documents and comparative lack of experience interacting with them as opposed to paper texts for even the most dedicated users. The lack of standards in the electronic domain compared to the rather traditional structures of many paper documents is a further problem for schema development with contemporary electronic texts.

2.9.2 Navigation: acquiring a cognitive map of the text

If picking up a new book can be compared to a stranger entering a new town (i.e., we know what each is like on the basis of previous experience and have expectancies of what we will find) how do we proceed to develop our map of the information space?

Current theories of how humans navigate vary and it is no longer the province of psychologists alone. Geographers, anthropologists and urban planners all show an interest (see for example Downs and Stea, 1974). However, Tolman's (1948) paper on cognitive maps is frequently cited as seminal. He postulated the existence of a cognitive map, internalised in the human mind which is the analog to the physical lay-out of the environment. In dismissing much of the then popular behaviourist school of psychology, Tolman argues that information impinging on the brain is:

> "worked over and elaborated....into a tentative cognitive-like map of the environment indicating routes and paths and environmental relationships.." (p. 192).

Recent experimental work takes the notion of some form of mental representation of the environment for granted, concerning itself more with how such maps are formed and manipulated. Many theorists agree that the acquisition of navigational knowledge proceeds through several developmental stages from the initial identification of landmarks in the environment to a fully formed mental map. One such developmental model has been discussed by Anderson (1980) and Wickens (1984) and is briefly described here.

According to this model, in the first instance we represent knowledge in terms of highly salient visual *landmarks* in the environment such as buildings, statues, etc. Thus we recognise our position in terms relative to these landmarks, e.g., our destination is near building X or if we see statue Y then we must be near the railway station and so forth. This

knowledge provides us with the skeletal framework on which we build our cognitive map.

The next stage of development is the acquisition of *route* knowledge which is characterised by the ability to navigate from point A to point B, using whatever landmark knowledge we have acquired to make decisions about when to turn left or right. With such knowledge we can provide others with effective route guidance, e.g., "Turn left at the traffic lights and continue on that road until you see the Bull's Head public house on your left and take the next right there..." and so forth. Though possessing route knowledge, a person may still not really know much about his environment. A route might be non-optimum or even totally wasteful.

The third stage involves the acquisition of *survey* knowledge. This is the fully developed cognitive map that Tolman (1948) described. It allows us to give directions or plan journeys along routes we have not directly travelled as well as describe relative locations of landmarks within an environment. It allows us to know the general direction of places, e.g., "westward" or "over there" rather than "left of the main road" or "to the right of the church". In other words it is based on a world frame of reference rather than an ego-centred one.

While such theoretical work on navigation is primarily concerned with travels through physical space such as cities and buildings it does offer a perspective that might prove insightful to the design of electronic, and particularly hypertext systems, where navigation is conceptualised as occurring through an information space. As reported earlier, navigation is considered to be the major process difference between reading from paper and from screen.

With texts, one would expect that generic structures such as indices, contents, chapter headings and summaries, page numbers and so forth be seen as landmarks that provide readers with information on where they are, just as signposts, buildings and street names aid navigation in physical environments. Thus when initially reading a text readers might notice that there are numerous figures and diagrams in certain sections, none in others, or that a very important point or detail is raised in a section containing a table of numerical values. In fact, readers often claim to possess such knowledge and there is some empirical evidence to suggest that this is, in fact, the case.

Rothkopf (1971) carried out an experiment to test whether such occurrences had a basis in reality rather than resulting from popular myth supported by chance success. He asked people to read a 12 page extract from a book with the intention of answering

questions on content afterwards. What subjects didn't realise was that they would be asked to recall the location of information in the text in terms of its occurrence both within the page (divided into eighths) and the complete text (divided into quarters). The results showed that incidental memory for locations within any page and within the text as a whole were more accurate than chance i.e., people could remember location information even though they were not asked to. There was also a positive correlation between location of information at the within-page level and accuracy of question answering.

There have been several follow-up studies by Rothkopf and other investigators. Zechmeister and McKillip (1972) had subjects read eight pages of text typed into blocks with four blocks per page. Subjects were asked to read the text before being tested on it. The test consisted of fill-in-the-blank questions, confidence ratings on their answers and location of the answer on the page. Again, an effect for knowledge of location was observed which was correlated to accuracy of answers, suggesting that memory for location and for content are independent attributes of memory that can be linked for mnemonic purposes. Interestingly no interaction of memory for location and confidence in answer was found. Further work by Zechmeister *et al.* (1975) and by Lovelace and Southall (1983) confirm the view that memory for spatial location within in body of text is reliable even if it is generally limited.[11]

Jones and Dumais (1986) empirically tested spatial memory over symbolic memory for application in the electronic domain, citing the work of Rothkopf and others as indicators that such memory might be important. In a series of three experiments they had subjects simulate filing and retrieval operations using name, location or a combination of both stimuli as cues. Like the preceding work on texts they found that memory for location is above chance but modest compared to memory for names and concluded that it may be of limited utility for object reference in the electronic domain.

Support for the notion of landmarks as a first level of navigational knowledge development in electronic text domains are provided by several studies which have required subjects to draw or form maps of the information space after exposure to it (see e.g., Simpson and McKnight 1990). Typically, subjects can group certain sections together but often have no idea where other parts go or what they are connected to. In other words, they possess a rudimentary mental map based on several base-points or landmarks and a few related items.

---

[11] In psychology, to describe a phenomenon as "reliable" implies that it is a non-chance occurrence. The implication in this context is that readers do indeed have memory for spatial location of information within a text but that it does not exist for all information: thus it is reliable but limited.

Unfortunately it is difficult to chart the development of navigational knowledge beyond this point. Detailed studies of users interacting with hypertext systems beyond single experimental tasks and gaining mastery over a hypertext document are thin on the ground. Edwards and Hardman (1989) claim that they found evidence for the development of survey type navigational knowledge in users exposed to a strictly hierarchical database of 50 screens for a single experimental session lasting, on average, less than 20 minutes. Unfortunately the data is not reported in sufficient detail to assess critically such a claim but it is possible that given the documents highly organised structure, comparatively small size and the familiarity of the subject area (leisure facilities in Edinburgh) such knowledge might have been observed.

In the paper domain the analogy with navigation is difficult to maintain at the survey and route levels as the reader can perform movements and gain direct access to all parts of the text in a manner that has no correlate in physical space. The information space in paper is immediately and directly accessible from all parts of the text and "taking a wrong turn" hardly has the consequences for the reader it does for the physical traveller relying on route knowledge. The analogy holds firmer with electronic space where constraints of manipulation and vision prevent the same access as paper.[12]

Obviously this is an area that needs further empirical work but as a source of potential difference between paper and screen, navigation would appear important. There are few reports of readers being lost with paper texts but many with electronic ones, from which one can only conclude that navigation will remain a concern of researchers for the foreseeable future. The following section describes some of the attempts that have been made to lessen it as a problem in the electronic domain.

2.9.3 Browsers, Maps and Structural Cues

A graphical browser is a representation of the structure of the database aimed at providing the user with an easy to understand map of what information is located where. According to Conklin (1987) graphical browsers are a feature of a "somewhat idealized hypertext system", recognising that not all existing systems utilise browsers but suggesting that they are desirable. The idea behind a browser is that the document can be

---

[12] On making this point once, a colleague retorted that movement via helicopter might be the physical equivalent of the direct access available to readers of paper texts. This does not seem appropriate to me as even with such a vehicle there are limitations on where you can travel (e.g., into buildings) which do not exist in paper texts and even the best designed vehicles fail to provide the speed of access to all parts of the environment that one's fingers offer with a well-bound book.

represented graphically in terms of the nodes of information and the links between them, and in some instances, that selecting a node in the browser would cause its information to be displayed.

It is not difficult to see why this might be useful. Like a map of a physical environment it shows the user what the overall information space is like, how it is linked together and consequently offers a means of moving from one information node to another. Indeed, Monk *et al.* (1988) have shown that even a static, non-interactive graphical representation is useful. However, for richly interconnected material or documents of a reasonable size and complexity, it is not possible to include everything in a single browser without the problem of presenting 'visual spaghetti' to the user. In such cases it is necessary to represent the structure in terms of levels of browsers, and at this point there is a danger that the user gets lost in the navigational support system!

Some simple variations in the form of maps or browsers have been investigated empirically. In a non-hypertext environment Billingsley (1982) had subjects select information from a database aided by an alphabetical list of selection numbers, a map of the database structure or no aid. The map proved superior, the no aid group performing worst.

In the hypertext domain a number of studies by Simpson (1990) have experimentally manipulated several variables to do with structural cues and position indicators. She had subjects perform a series of tasks on articles about houseplants and herbs. In one experiment she found that a hierarchical contents list was superior to an alphabetic index and concluded that users are able to use cues from the structural representation to form maps of the document. In a second study she reported that users provided with a graphical contents list showing the relationship between various parts of the text performed better than users who only had access to a textual list. Making the contents lists interactive (i.e. selectable by pointing) also increased navigational efficiency.

Manipulating "last card seen" markers produced mixed results. It might be expected that such a cue would be advantageous to all users but Simpson reported that this cue seemed of benefit only during initial familiarisation periods and for users of non-interactive contents lists. Further experiments revealed that giving users a record of the items they had seen aided navigation much as would be expected from the literature on physical navigation which assumes that knowledge of current position is built on knowledge of how you arrived there (Canter 1984). In general, Simpson found that as accuracy of performance increased so did subjects' ability to construct accurate post-

task maps of the information space using cards.

Such work is important to designers of hypertext systems. It represents a useful series of investigations into how "contents pages" for hypertext documents should be designed. Admittedly, it concerned limited tasks in a small information space but such studies are building blocks for a fuller understanding of the important issues in designing hypertext systems.

2.9.4 The provision of metaphors

A metaphor provides a way of conceptualising an object or environment and in the information technology domain is frequently discussed as a means for aiding novices comprehension of a system or application. The most common metaphor in use is the desk-top metaphor familiar to users of the Apple Macintosh amongst others. Here, the user is presented with a virtual desktop on screen and can perform routine file manipulations by opening and closing "folders" and "documents" and throwing them in the "wastepaper bin" to delete them. Prior to this metaphor, the word processor was often conceptualised by first-time users as a typewriter.[13]

The logic behind metaphors is that they enable users to draw on existing world knowledge to act on the electronic domain. As Carroll and Thomas (1982) point out:

> "If people employ metaphors in learning about computing systems, the designers of those systems should anticipate and support likely metaphorical constructions to increase the ease of learning and using the system" (p.108).

However, rather than anticipate likely metaphorical constructions, the general approach in the domain of hypertext has been to provide a metaphor and hope (or examine the extent to which) the user can employ it. As the term 'navigation' suggests, the most commonly provided metaphor is that of travel.

Hammond and Allinson (1987) report on a study in which two different forms of the travel metaphor were employed: "go-it-alone" travel, and the "guided tour". These two forms were intended to represent different loci of control over movement through the document, the first being largely user-controlled and the second being largely system-controlled. Additionally a map of the local part of the information structure was available

---

[13] The history of technological progress is littered with such metaphors e.g., the car as a "horseless carriage", the first typefaces were imitations of script and so on.

from every screen, with selectable arrows at the four edges leading to further maps, frames so far visited indicated, and all frames directly selectable from the map. Hammond and Allinson stress the importance of integrating the metaphor in the design of the system, which they did, and not surprisingly they found that users were able to employ it with little difficulty.

Of course, one could simply make the electronic book look as similar to the paper book as possible. This is the approach advocated by people such as Benest (1989) with his book emulator and as such seems to offer a simple conceptual aid to novice users. Two pages are displayed at a time and relative position within the text can be assessed by the thickness of pages either side which are splayed out rather like an opened paper document. Page turning involves a single mouse press which results in two new pages appearing or by holding the mouse button down and simulating "flicking" through the text. The layout of typical books can also be supported by such a system thereby exploiting the schematic representations we know that experienced readers possess.

If that was all such a system offered it would be unlikely to succeed. It would just be a second-rate book suffering from basic ergonomic and manipulation problems outlined earlier. However, according to Benest, his book emulator provides added-value that exploits the technology underlying it. For example, although references in the text are listed fully at the back of the book they can be individually accessed by pointing at them when they occur on screen. Page numbers in contents and index sections are also selectable thereby offering immediate access to particular portions of the text. Such advantages are typical of most hypertext applications. In his own words:

> "the book presentation, with all the engrained (*sic*) expectations that it arouses and the simplicity with which it may be navigated, is both visually appealing and less disruptive during information acquisition, than the older 'new medium demands a new approach' techniques that have so far been adopted" (p.63).

This may be true but Benest has offered no supporting evidence and in the absence of empirical data one should view all claims about hypertext with caution.

It is interesting for two reasons that Benest dismisses the "new medium demands a new approach" philosophy of most hypertext theorists. Firstly, there is a good case to be made for book-type emulations according to the arguments put forward above about schematic representations readers possess of texts. As outlined earlier, such representations facilitate usage by providing orientation or frames of reference for naïve

users. Such points have been raised in sufficient detail earlier to not require further elaboration here. Secondly, the new approach which rejects such emulations has largely been responsible for the adoption of the concept of navigation through electronic space.

In response to the first issue it is worth noting that Benest's approach is, probably correct up to a point. There are attributes of paper texts that are worth retaining for the new medium. However, retention of useful structures is not to be equated with emulation of paper. Retention does not extend as far as mimicking page-turning or providing splayed images of the pages underlying either opened leaf. Furthermore while identifying relevant schematic structures for texts is advisable, one would not expect all text types to retain such detailed aspects of their paper versions in hypertext. There seems little need, for example, to emulate the book form to this degree for a hypertext telephone directory. Benest does not seem to draw the line however between texts that might usefully exploit such emulations and those that would not, or state what he would expect unique hypertext documents to emulate.

In response to the second point, it is worth asking is there an alternative to navigation as a metaphor? As noted in this chapter, the dominant approach to hypertext has produced the navigation through space metaphor. Benest, though still talking of navigation, does so in the limited sense that it is used in the paper domain. The more typical hypertext approach embraces navigation whole-heartedly and uses it as a means of inducing orienting schemata in the users mind.

Hammond and Allinson (1987) discuss the merits of the metaphor approach in general and the navigation one in particular for hypertext. They argue that there are two relevant dimensions for understanding the information metaphors convey: *scope* and *level* of description. A metaphor's scope refers to the number of concepts that the metaphor relates to. A metaphor of broad scope in the domain of HCI is the desk-top metaphor common to many computing interfaces. Here, many of the concepts a user deals with when working on the system can be easily dealt with cognitively in terms of physical desk-top manipulations. The typewriter metaphor frequently invoked for explaining word processors is far more limited in scope. It offers a basic orientation to using word processors (i.e., you can use them to create print quality documents) but is severely limited beyond that as word processors do not behave like typewriters in many instances.

The metaphor's level of description refers to the type of knowledge they are intended to convey. This may be very high level information such as how to think about the task and its completion, or very low, such as how to think about particular command syntax in

order to best remember it. Hammond and Allinson talk of four levels: task, semantic, lexical and physical which refer to general issues such as: "Can I do it?"; "What does this command do?"; "What does that term mean?" and "What activities are needed to achieve that?" respectively. Few, if any, metaphors convey information at all levels but this does not prevent them being useful to users. In fact, few users ever expect metaphors to offer full scope and levels of description.

According to Hammond and Allinson the navigation metaphor is useful in the hypertext domain and when users are offered "guided tours" through an information space they do not expect physical manifestations of the metaphor to apply literally but might rely primarily on semantic mappings between metaphor and system much more heavily. As there are numerous rich mappings that can be made between the navigation metaphor and hypertext it seems sensible to use it.

Benest's book emulation is also a metaphor for using the system and in some instances would offer a broad scope and many levels of description between the paper text and the hypertext. The fact that one can talk about navigation and book metaphors in the one system shows that mixed metaphors are possible and (though awaiting confirmatory evidence) probably workable in some instances.

It is hard to see any other metaphors being employed in this domain. Navigation is firmly entrenched as a metaphor for discussing hypertext use and book comparisons are unavoidable in a technology aimed at supporting many of the tasks performed with paper documentation. Whether there are other metaphors that can be usefully employed is debatable. Limited metaphors for explaining computer use to the novice user are bound to exist and where such users find themselves working with hypertext new metaphors might find their way into the domain. But for now at least it seems that navigation and book emulation are here to stay.

2.9.5 Conclusions

Information structure–related issues are currently the major research area in the electronic text domain. Borrowing heavily from mainstream cognitive psychology, concepts such as schemata, models and mental maps are seen now as pertinent to the analysis of reading from screens.

Readers do seem to possess models of a text's typical form and this has direct relevance to the creation of electronic versions. For the transfer of existing paper formats to

hypertext it is likely that the maintenance of superstuctural cues is important. For the creation of original electronic texts it is important to provide suitable means for schematic abstractions on the part of the reader. Quite how this might be done is yet to be resolved.

Existing theories of navigation in physical space have some application in the electronic text domain although the mapping is rarely direct. It is more usual to talk of navigation as if it was a problem unique to electronic information systems. Work on providing cues has been productive, and some guidelines on the design of maps and contents representations are emerging. With respect to metaphor provision it is unlikely that the navigation and book metaphors will be replaced though the empirical evidence for their use or optimum presentation is unfortunately lacking.

The major shortcoming of much of the literature on these themes is that it is discursive rather than empirical. The dominant philosophy underpinning much of it is that hypertext is intrinsically better than other text forms and that technology has at last provided us with the opportunity to design information sources suitable for man's cognitive style (however that is defined). This may prove to be the case, but in the absence of a sound empirical base there are likely to be many mistakes made before optimum designs are derived. As the solution to the problems inherent in reading from screens, hypertext has yet to prove itself.

2.10 General conclusions and the way forward

A large and diverse literature on electronic text now exists which is added to regularly by theorists and practitioners from a variety of disciplines such as ergonomics, psychology, typography, information science and computer science. Making sense of it all is a difficult task but conceptualisation can be aided by considering the readers, their tasks and the information space as the important variables to understand.

From the work reviewed here it is clear that the issues raised in reading from screens alter as a function of the type of reading involved. The image quality hypothesis of Gould *et al.* for example is applicable to proofreading short texts, when the dependent variables are speed and accuracy. However, this "solution" has little application for larger texts when associated issues of manipulation become involved and speed is less important or for situations where the reader must locate an argument in a book length text and cannot formulate a specific search string.

The classification proposed in Figure 2.1 aims to account for such variation by grouping the issues in terms of size of information space, breadth of issue covered and specificity of prediction that can be made from current knowledge. This classification makes no claims to completeness but does seek to reflect the major issues of concern from the user's or reader's point of view and as such, provide a means of interpreting and making sense of this expanding literature.

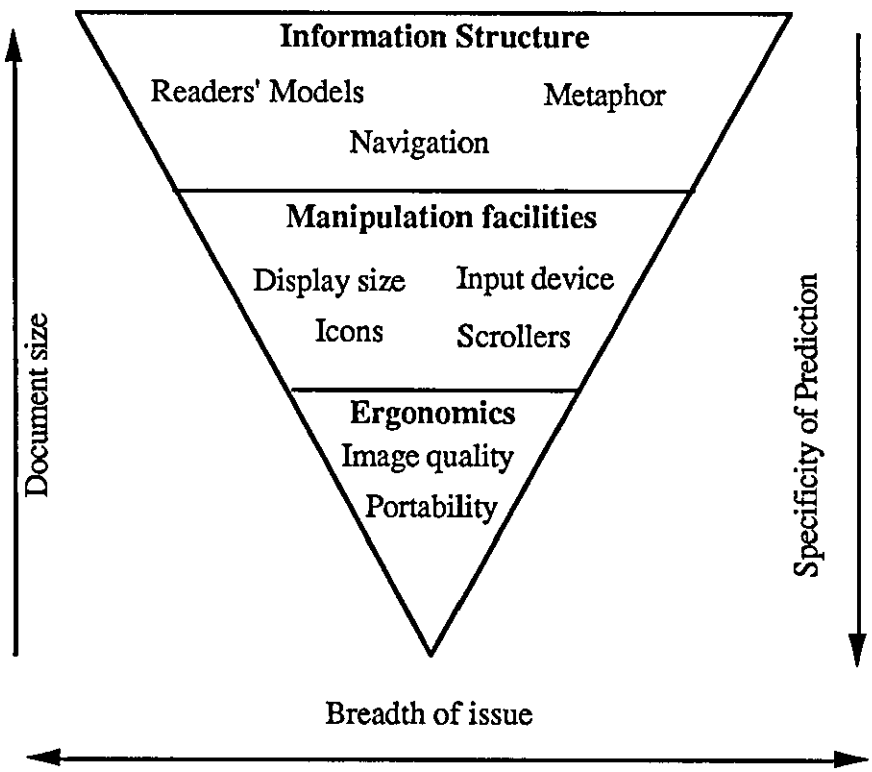


Figure 2.1. A classification scheme for research on reading electronic text

This is a three-tier classification representing the trends identified in the research over the last 15 years. As mentioned before it loosely equates to a distinction between perceptual, motor and cognitive issues on the part of the reader. At the simplest level, reading a single screen or page of text (such as a note or letter), the major limiting factors are perceptual, the basic or visual ergonomics. Once the information space becomes larger than a single screenful the user must manipulate the information in order to view it all. This necessitates the use of whatever facilities are available in the interface, requiring motor behaviour on behalf of the reader and focuses attention on aspects such as button selection, scrolling, searching mechanisms etc. As the information space becomes even larger it seems that more cognitive issues such as the

reader's model of its structure start to play an increasingly important part, aiding navigation, suggesting where information might be found and so forth.[14]

Negatively correlated with size of information space is the specificity of the prediction that can be made about electronic text usage on the basis of current knowledge. From the basic ergonomic research it is now possible to predict accurately the likely effect of image quality on reading a short text from screen. The 20-30% speed decrement for typical screen reading is a robust finding (though typical screens should improve over time thereby lessening this effect). At the other end of the spectrum, for lengthy texts, such predictions cannot be as precisely made.[15]

Positively correlated with size of information space is the breadth and range of issues involved. The factors influencing reading of lengthy texts are broader and more numerous than those influencing short text. The psychological issues involved move away from the mainly perceptual or visual (as claimed by Gould *et al.*, 1987a, 1987b) to the mainly cognitive and encompass many of the areas of concern to psychologists that led Huey (1908) to state that explaining reading would be the acme of any psychologist's achievements.

Perhaps the most important conclusion to draw from this literature is that ultimately, the explanations of differences and means for improvement lie not at one level exclusively but in the interaction of all three. Good image quality or cognitively compatible structures alone will not ensure usability but the combination of these with suitable manipulation facilities should at least minimise the differences between the two media and for some tasks and texts even convey advantages to the electronic medium.

There are situations where one level of issue is obviously more important than the others, e.g., when presenting a very short text on screen such as a letter or memo, manipulation and information structure research is of little applicability and image quality findings take precedence. On the other hand, simply addressing image quality will not ensure the design of a usable electronic software manual for example. Only by focussing on the particulars of the reading task can real insight be gained.

---

[14] The mapping between the literature as reported here and the perceptual, motor and cognitive aspects of reading is not as direct as it superficially appears however. Portability for example is a basic ergonomic issue but could not be described as a perceptual one as would be implied if the mapping was direct.

[15] For certain tasks such as locating material in a book-length text, it is possible to predict that search facilities convey advantages to the electronic medium unmatched by paper. However such predictions do not invoke the higher-level or cognitive concepts of schemata or navigation and are not therefore specific predictions derived from work on information structuring.

The importance of task variables cannot be overstated. Without specifying the type of reading being referred to, it is impossible to extract relevant findings from the literature. The notion of task is implicit in the classification proposed here which emphasises the importance of document size and breadth of issue. Unfortunately much of the literature is vague on task matters and makes sweeping generalisations from the particulars of one study to the generalities of reading from screens.

This is often the case in much of the basic ergonomic work on speed and accuracy of reading short texts. Such work only informs us of reading short texts in a particular way with a particular goal in mind. Yet, researchers reporting such work frequently conclude that on the basis of their findings "reading from screens is/is not faster/slower etc. than paper" and so forth without any reference to task effects. The attitude is neatly epitomised by Oborne and Holton (1988) who published an account of one such study under the ambitiously sweeping title: "Reading from Paper versus Screen: there is no difference"!

The variation of texts used in many of these studies is also a cause for some concern. All texts are not the same and reporting no outcome differences for a specified task on a 1500 word text does not mean that all 1500 word texts can or will be read to a similar standard for an equivalent task. The most common experimental texts are descriptive extracts from paper documents, often ones of no particular interest or relevance to the subjects. This hardly reflects the type of reading routinely performed by people everyday and is a major shortcoming of much of the work on reading from screens.

In sum, research suggests that reading from screens is a complex issue for which simple recommendations to designers cannot be made. The exact role of electronic text in the information age has yet to be defined but it is clear that simply reproducing paper versions of texts electronically will not work. Marshalling the findings of various researchers into a cohesive body of knowledge is hampered by the lack of a descriptive framework with which to discuss readers, texts and tasks. Each discipline has its own conceptualisation of the reading process and its outcomes with the result that the literature is fragmentary and broad. The present thesis can be seen as an attempt to rectify this shortcoming by developing a descriptive framework that facilitates usable electronic text design.

The present review aids this quest by identifying the crucial issues that electronic text designers must address. The three levels of issue provided in the above classification

offer a clue to the aspects of reading that are important and which therefore must be considered in design. As a result, the framework under development in this work needs to encapsulate these in order to ensure that they are not overlooked when an electronic text application is being specified. A typical applications designer cannot be expected to have a detailed knowledge of the literature. A suitable framework should overcome this deficit, at least in part by informing him of the most pertinent usability factors and providing an orientation towards the design problem that leads to the formulation of appropriate queries of the knowledge base. Furthermore, detailed knowledge of the literature can act as a test of the validity of any proposed framework. If it is to be useful it should support the drawing of conclusions about reader performance with an electronic text that are broadly in line with experimental findings. This aspect will be considered further later in the thesis.

In conclusion it can be said that prescriptive guidance on the design of electronic texts is in short supply in the published literature on reading from screens. However there are several levels of issue that any designer needs to consider which consequently need to be included in any tool, be it a framework, a model or a set of guidelines aimed at supporting the designer. In the following chapter consideration is given to what else might need to be included in the descriptive framework that is the goal of this thesis and the level of detail at which it might best be presented.

CHAPTER 3

DESCRIBING THE READING PROCESS: THE NEED FOR AN APPROPRIATE LEVEL

3.1 Introduction

Taken in isolation, the literature on reading from screens versus paper is largely unhelpful to a designer concerned with developing an electronic text system. The classification proposed in the previous chapter is an attempt to afford better conceptualisation of the relevant issues and experimental findings but it suffers from the problems inherent in the literature itself: the absence of a suitable descriptive framework of the reader that would enable those concerned with electronic text to derive guidance for specific design applications.

As it stands, the literature presents two implicit views of the typical reader and provides recommendations accordingly. The first is as a scanner of short texts, searching out spelling mistakes or some such trivial error. The second is as a navigator through a maze of information. These extremes are tempered only slightly by concessions to text or task variables as influences on the reading process and it is rare that any attempt to place reading in a broader, more realistic context is made. Yet reading rarely occurs as a self-contained process with goals expressible in terms of speed or number of items located. Far more frequently it occurs as a means to, or in support of, non-trivial everyday ends such as keeping informed of developments at work, checking bank balances, understanding how something works and so forth. These are precisely the type of acts that people perform routinely, of which reading is an essential component. Their success is measured in effects not usually quantifiable in time or error scores. True, readers do notice spelling mistakes, they may even proofread as professionals and they certainly must navigate lengthy documents, but such views alone can never adequately describe the realities and totality of the reading situation.

The constrained view of the reading process becomes even more apparent when one examines the conceptualisations of reading that dominate the various disciplines which lay claim to some interest in it. Psychology, a discipline that might justifiably consider itself directly concerned with understanding reading is, according to Samuels and Kamil (1984), concerned with:

> "the *entire* process from the time the eye meets the page until the reader experiences the 'click of comprehension'" (p. 185, italics added).

This sounds suitably all-embracing but in reality is relatively narrow when one realises the everyday attributes of reading that it overlooks. There are few psychological models of reading that consider text manipulation or navigation for example as part of the reading process. The literature which provides theoretical input to these domains is usually the product of other research issues such as memory organisation and learning. Furthermore not all psychologists would even accept Samuels and Kamil's definition of psychology's legitimate concerns with reading. Crowder (1982) for example, explicitly excludes comprehension as an issue for psychological investigation in his analysis of reading.

A typical psychological theory of reading is Rumelhart's (1977) which marked a supposed breakthrough in cognitive models of reading by highlighting the limits of linear models (those proposing a one-way sequence of processing stages from visual input to comprehension). He outlined an alternative interactive form (one supporting the influence of higher stages of processing on lower ones) which accounted for some of the experimental findings that were difficult to accommodate in linear models. His model, parts of which have been successfully implemented in software form (McClelland and Rumelhart 1981), represents reading largely as an act of word recognition. The model has been summarised as follows:

> "Reading begins with the recognition of visual features in letter arrays. A short lived iconic image is created in brief sensory storage and scanned for critical determinants. Available features are fed as oriented line segments into a pattern synthesizer that, as soon as it is confident about what image has been detected, outputs an abstract characterisation.....The extracted features are constraints rather than determiners, interacting with context and reader expectations....The individual letters are heavily anticipated by stored representations in a 'word index'. Even in recognition of letters and words all of the various sources of knowledge, both sensory and non-sensory, come together in one place and the reading process is the product of the simultaneous joint application of all knowledge sources." (de Beaugrande, 1981, p. 281).

From such a description it is not difficult to understand why Venezky (1984) states that:

> "the history of research on the reading process is for the most part the history of cognitive psychology," (p. 4)

but it strongly emphasises the limitations of such work for system designers. There is no mention of a reading task, a text, a goal, or a context in which these processes occur. 'Visual feature recognition', 'iconic images' and 'pattern synthesizers' are theoretical constructs which attempt to provide a plausible account of how humans extract information from text (or 'letter arrays' to use the jargon) but mapping findings from such analyses and models to the world of electronic text design would seem to be beyond the model's scope.

The overly narrow focus of much reading research is reflected sharply in the opening pages of psycholinguist Frank Smith's (1978) book titled *Reading* where he remarks that a glance though the text might leave one justifiably thinking that:

> "despite its title, this book contains little that is specifically about reading" (p. 1).

He goes on to argue for the need to examine reading in a reductionist manner (claiming that there is little about reading that is unique – it involves certain cognitive processes and structures that researchers not interested in reading have already investigated in other contexts – presuming that such work transfers appropriately to discussions of the reading process which is itself a debatable assumption), but it is his early admission of the lack of real-life relevance of the work that stays longest in mind.

The most recent and vociferous attack on the standard cognitive psychological approach to reading comes from Kline (1988). In a book given over to attacking much of experimental psychology he singles out reading as a prime example of the lack of ecological validity in much of the discipline's work. Describing a typical reading experiment investigating people's categorisation of sentences as meaningful or non-meaningful while their reaction time is measured, he states:

> "the task ... is not really like reading. Certainly it involves reading but most people read for pleasure or to gain information. Furthermore, reading has serious emotional connotations on occasion, as for example reading pornography, a letter informing you that your PhD has been turned down (!) or your lover is pregnant ..... Furthermore, most adults, when reading books especially, read large chunks at a time" (p. 36).

He continues, humorously comparing lines from a Shakespeare sonnet (e.g.: Like as the waves make towards the shore, so do our minutes hasten to their end...) with lines

from such experimental tasks (e.g.: Canaries have wings - true or false?; Canaries have gills - true or false?) before concluding that such work is absurd in the context of real reading and its resultant theoretical models of no predictive and little explanatory use. His criticisms might seem harsh and populist were it not for the fact that Kline is a psychologist of international reputation and admits to having performed such experiments himself earlier in his career.

But psychology is not unique in failing to provide a satisfactory account of the process. Information science, the theoretical backbone of librarianship might also be viewed as having a natural interest in the reading process. Yet its literature offers few clues to those concerned with designing electronic texts for reading. As Hatt (1976) puts it:

> "A great body of professional expertise has been developed, the general aims of which have been to improve the provision of books and to facilitate readers' access to books. At the point where the reader and the book come together however, it has been the librarian's habit to leave the happy pair and tiptoe quietly away, like a Victorian novelist." (p.3).

Hatt argues that the problems of information science are the problems of all disciplines concerned with this subject and that although much valuable work has been done and knowledge has been gained he draws a similar (though less emphatically expressed) conclusion to Kline in that one comes away from the literature thinking "that's all well and good, but it's not really reading!"

In defence of each discipline it must be said that their approaches reflect their aims. If psychology really is concerned with what happens between the moments when the eye meets the page and the reader understands the text (or just before that in the case of Crowder *inter alia*), then models of eye movements and word recognition have a place, despite Kline's enthusiastic dismissal.[1] Few, if any, theorists interested in reading claim to cover all issues. What is pertinent here however is the irrelevance of much of this work to the issues associated with electronic text.

The unsuitability of any theoretical description of reading is a major problem for human factors work. Viewing the reader as an "an information processor" or "a library user" and the reading process as a "psycholinguistic guessing game" depending on theoretical stance hardly affords prescriptive measures for the design of electronic text systems.

---

[1] Though one might add that if this is all many psychologists consider important in understanding the reading process then Kline might really have a point.

The reader does process information and occasionally uses libraries, but each is only a small part of the whole that is reading. If one deals exclusively with such aspects as many theories do, the broad picture never emerges and this gives rise to the type of limited findings on text design one finds in the human factors literature.

3.2 The problem of theoretical description for human factors work

The problem for human factors work induced by inappropriate descriptions is epitomised in a case study of a commercial system involving the author (Dillon 1988 - see Appendix A). A publishing consortium recently funded the development of an experimental system to support the document supply industry. Named ADONIS, the resulting workstation is designed to facilitate searching, viewing and printing of CD-ROM stored articles. It boasts a high-resolution A4-size screen that presents bit-mapped reproductions of journal articles. The trial system presented users with access to biomedical journals (selected on the basis of a usage study) and the workstation was aimed primarily at document supply staff working in storage centres who would process inter-library loan requests for articles, though possibilities for end-users in libraries to use the system directly existed and were considered by the development consortium as worthy of investigation. Thus the system could be seen as a prototype electronic text system of the future.

The author was asked to evaluate the system from a human factors perspective for both user types.[2] ADONIS' application to the document supply industry was a relatively straightforward evaluation and will not be discussed further here except to emphasise that it became clear from the task and user analyses carried out on site that these users did not view documents with the system and under no circumstances could be described as readers of the material contained in the database. The potential end-users in libraries on the other hand were typical readers of journal articles and the evaluation in this context highlighted the shortcomings in current knowledge of electronic text design.

The specific details of the evaluation are not important for the present discussion. Time allowed for this work was limited as the developers were not keen on lending the equipment to remote sites for extended periods of time thus necessitating a quick walkthrough rather than formal experimental approach on the part of the author.

---

[2] The complete system was subject to a full assessment by a team of researchers at HUSAT consisting of the author and two others. However the work was conveniently divided amongst the team and the usability evaluation reported here was solely the work of the present author.

Suffice to say that the evaluation involved setting three tasks for a sample of 10 users to perform in an informal manner i.e., with the evaluator present and the participants commenting on the system's user interface as they worked through the tasks. The tasks were so designed to ensure that subjects were exposed to all aspects of the interface. Measures of speed and accuracy were eschewed in favour of general ratings of the system and comments on good or bad aspects of the interface which the evaluator noted as the subject proceeded. As a result of frequent criticisms of the search facilities, a survey of normal procedures for citing articles amongst 35 researchers was also carried out. The results of the evaluations were summarised, related to the literature on electronic text and general interface design and presented to the publishing consortium (in the form of the report in Appendix A).

On the surface, ADONIS was a good design. The high quality screen presented articles in an easy to read manner that conformed precisely to the structure of the paper version. By using a large screen and positive presentation it even adhered to some of the human factors design guidelines in the literature. Use of menus and a form filling screen for inputting search criteria should have removed any learning difficulties for novice users too. The ability to store and retrieve a large number of articles from one system coupled with the ability to view material on screen before deciding whether or not to print it out, would seem to convey benefits to the ADONIS workstation not even possessed by paper, never mind other databases.

The users studied at HUSAT however were very critical of the system. Common criticisms related to the rather "archaic" style of the ADONIS interface, the speed of searching which was perceived as far too slow, the inability to search on keywords and the restricted manipulation facilities available once an article was being viewed (ADONIS by virtue of its reliance on bit-mapped images was slow and only supported paging forward or backwards, jumping directly to particular sections was impossible).

In an attempt to understand the likelihood of potential users actually reading texts with ADONIS, they were asked to comment on the readability of the displayed document independently of the manipulation facilities offered. Only two users said they would read with it; of the remainder, six said they would only scan articles prior to printing them out and two said they would never use it. In other words, although the system was designed in partial accordance with the literature on electronic text, it was rejected by users. How can this be?

What was shown by the evaluation is that while ADONIS supports the ends, it fails to

adequately provide the means. In other words, it will let users get *what* they want but not *how* they want to. Users can obtain hardcopies of journal articles but they must master the counter-intuitive specification form first.[3] They can browse articles on a high quality large screen, but they cannot manipulate pages with ease as they do with paper. They can search an equivalent of a library of journals from their desk to obtain an article they seek, but they cannot browse through a list of contents and serendipitously discover a relevant title or author as they can with paper.

This clash between means and ends provides an interesting insight into the problems faced by many designers of electronic text systems (or indeed, information systems in general) which will be referred to here as the "levels of description" problem. Briefly stated, it implies that there are various levels of abstraction at which human behaviour can be described and while each may be accurate in itself there exists an optimum level for any given use (e.g., analysing consumer spending requires different views of human activity than describing human task performance when driving). In the case of systems design, using a non-optimum level leads to superficial matching of design to needs if the level is too shallow, and to an inability to specify designs for needs if the level is too deep. These will be elaborated with two examples pertinent to electronic text design.

ADONIS seems to match basic reader requirements. However, it is obvious that it does so only at a superficial level. By describing readers' needs at the gross level of "obtaining a hardcopy", "locating an article", "browsing the references" and so forth it has made (and matched) design targets of accurate but inadequately specified needs. The designers obviously developed a product that satisfies these needs, but only at the gross level of behaviour. A description of reading at a deeper level than this might well have produced a different set of requirements and resulted in a more usable design.

An example of a level of description too deep to specify needs for design can be found in most of the work on modelling reading by cognitive psychologists. By concentrating on theoretical structures and processes in the reader's mind or eye movements in sentence perception, word recognition and so forth, such work aims to build a body of knowledge on the mental activities of the reader. Fascinating as this is, it is difficult to translate work from this level of description or analysis to the specification of software

---

[3] The survey of citation style revealed that users tend to refer to articles in the form author/year/title/journal, or author/ journal/article/year. ADONIS structured input in the form: ADONIS number/ISSN Number/journal/year/author etc. which was considered very confusing by some users and led to frequent errors during trial tasks (see Appendix A for further details).

intended to support the reading process.[4]

Highlighting limitations is only useful where it serves to advance the means of overcoming them. What is required therefore is a some level of discourse that bridges between these two extremes and actually provides valid descriptions of human activities in a form that leads to specific recommendations for system design. This is not an easy task but one is helped by at least knowing where the goalposts are. Within the electronic text domain a suitable analytic framework should provide designers with a means of posing appropriate questions and deriving relevant answers. Clearly existing ones, be they psychological, typographical or information science, do not. How then should we conceptualise the reading process? It is this question that is addressed in the following section.

3.3 Identifying an appropriate abstraction

It is unlikely that the evolution of a suitable description of the reading process will result merely from performing more experiments on reading from screens. To attempt empirical testing of all conceivable reading scenarios would be impossible and as has been shown by the ADONIS analysis, even the application of demonstrable ergonomic principles derived from such work (e.g., the importance of image quality) is insufficient to guarantee successful design.

For the purposes of designing systems for process control Rasmussen (1986) describes the need for a multi-layered analysis involving descriptions ranging from the social function of a system, through the information processing capabilities of the user and machine, to the physical mechanisms and anatomy of both the user and the machinery. He emphasises the need to incorporate perspectives of human abilities from quite separate research paradigms in order to describe usefully the process of interaction with advanced technology and adds:

> "it is important to identify manageable categories of human information processes at a level that is independent of the underlying psychological mechanisms" (p.99).

In other words, the framework needed for design should not be overly concerned with

---

[4]For a clear example of this see Rumelhart's (1977) widely acclaimed work on the development of a reading model (described earlier) and attempt to draw a set of guidelines from this that is applicable to HCI.

the architecture of human cognition (as is the case with most cognitive models of reading). Thus, according to Rasmussen, advances can be made on the basis of understanding the relevance of human information processing components (e.g., working memory, schemata etc.) without specifying their underlying structural form (e.g., as production systems (Anderson 1983), blackboard architectures (Hayes-Roth 1983) and so forth).

In HCI, the most popular behaviours to examine deal with text-editing, a task so heavily studied and modelled that it has earned the derogatory title in some quarters of the "white rat of human factors".[5] It is easy to see from such models that the Rasmussen approach of multi-layered, architecture-independent analysis is largely ignored. For example, one popular model of this activity, judged by citations rather than actual use in design, is based on the cognitive complexity theory (CCT) of Kieras and Polson (1985). This theory not only formally advocates the production system architecture of human cognition as a means of "calculating" learning difficulties in transferring between text editors, it addresses only one level of activity for the system, that of correcting spelling mistakes in previously created text. The accuracy of the model is often held up as an example to other researchers and theorists in HCI, even though its utility to designers remains, five years on, to be convincingly demonstrated.

Predictive modelling techniques for HCI rely on identifying small units of behaviour, decomposing them into their assumed cognitive primitives, analysing them with respect to time and errors, and then developing an approximate model which accounts for performance within certain boundaries such as error-free expert performance. Such models of user behaviour with technology exist not only for text editing but in less extreme forms for menu navigation (Norman and Chen 1988), item selection with input devices (Card *et al.*, 1978), so why not reading?

The crucial point is that reading could be equivalently modelled if ergonomists were to conceptualise it as narrowly as proofreading or item selection from a list of words. Indeed, models of such activities are beginning to emerge in the HCI domain (see e.g. Wright and Lickorish 1984). In equating reading with such activities complexity is certainly reduced but range of application is surely curtailed. Accurate models of

---

[5] This descriptor was first seen by the author in an item on HICOM, the electronic conferencing system of the British human factors community. The implication of the description is surely that text-editing tells us as much about human-computer interaction as a rat's performance tells us about being human. Depending on theoretical perspective that might mean a lot or a little, but given the absence of hardline behaviourists in HCI research one can only conclude that it is a little.

proofreading might eventually lead to prescriptive principles for designing screen layout and manipulation facilities for such tasks, rather like the GOMS model (Card *et al.*, 1983) can theoretically aid the design of command languages for systems, but they are unlikely to prove extensible to the wider issues of electronic text design such as what makes a good or bad electronic text, or how should a hypertext be structured?

There is a school of thought that suggests that while such questions cannot be answered yet the modelling approach is "good science" and that sufficient progress in applied psychology will eventually be made by the accumulation and refinement of such low-level, predominantly mathematics-based models. Newell and Card (1985) argue that in all disciplines, hard science (i.e, technical, mathematical) drives out the soft and that quantification always overpowers qualification. With reference to HCI they argue that psychology's proper role is to use its strengths and leave behaviour outside its remit to other disciplines. For these authors the domain of psychology covers human actions within the time scale of 0.1 to 10 seconds. Anything smaller, they claim, is covered by the laws of physics, chemistry and biology, anything larger than this but less than a matter of days is covered by the principles of bounded rationality and the largest time frame of weeks to years is the proper subject matter of social and organisational theories.

Within the narrow time scale they allow psychology, they propose that psychologists concentrate on "symbolic processing", "cycle times", "mental mechanics" (whatever they are!) and "short-term/long-term memory". They accept that the bounded rationality time-band covers many of the aspects of human behaviour relevant to HCI (and to humans in general it might be added) but

> "their theoretical explanation is to be found in the interplay of the limited processing mechanisms of the psychological band and the user's intendedly rational endeavours" (p.227).

Claims that this is too low level to be relevant to designers are partially correct, say the authors, but this problem will be overcome when a suitably all embracing model of human information processing has been developed that can be applied wholesale to issues at the level of bounded rationality, i.e., the level at which human activity is normally described and understood e.g., reading a book, writing a letter, driving a car and so forth.

This approach has been the subject of harsh criticisms (see e.g., Carroll and Campbell

1986) and contradictory evidence. On the basis of interviewing designers of videotext systems, Buckley (1989) concluded that the type of model proposed by Card *et al.* (1983) was irrelevant to most designers. He claims that designers tend to avoid academic literature and specific experimental findings in favour of their own internalised views of typical users and good design practice. Buckley states:

> "The designers expressed concern about the ease of use of the dialogues and had some clear views of how system features under their control would affect users.... But they did not report any use of traditional forms of human factors information which are expressions of the science base and are normally represented in research papers and design handbooks" (p.183).

Instead, he found that designers relied heavily on "pre-existing internalised frameworks" (p.184) which consist of primitive and weakly articulated models of users and their tasks that the system must support. Buckley goes on to emphasise the importance of providing information to designers in a form compatible with this style of working. Such findings are not unique, similarly doubtful views of the validity of formal models and standard human factors literature based on empirical findings have been expressed by other researchers who have interviewed designers (see e.g., Hammond *et al.* (1983) and Gardiner and Christie (1987).

This partly confirms the conclusions drawn from the ADONIS study where it was obvious that designers had some ideas of the users they were designing for, except that in this case, they were obviously also aware of some of the recommendations from the literature.[6] Regardless of their familiarity with the literature though, designers seem to have an idea of who their designing for and what tasks the system will support, how else could they proceed? Their views are naturally partial and often intuitive. Therefore making this conceptualisation more explicit and psychologically more valid in an appropriate way would seem to be of great potential benefit to the design world.

The second major flaw in Newell and Card's argument is that it assumes the world can afford to wait for an all-embracing cognitive model to emerge while all around us, technological advancement accelerates. They counter this criticism with the somewhat surprising statement that technology does not advance as fast as we think it does, but they are in a very small minority if they really believe this. Regardless of the level of advancement, in the domain of reading at least, cognitive psychological models of the

---

[6] This contrasts sharply with Buckley's subjects, some of whom registered surprise when he told them such a literature actually existed.

process exist which satisfy many of the criteria of hard, quantitative science (e.g. Just and Carpenter 1980) but as has been repeatedly pointed out, these just do not seem to afford much in the way of design guidance.

What seems to be required is a descriptive level above the information processing models advocated by Newell and Card but below the very high level descriptions of the bounded rationality approach favoured by information scientists. This is the level of Rasmussen's architecture-independent frameworks.

In the case of reading and electronic text systems a suitably embracing framework would need to cover the range of issues from why an individual reads to how the screen can be best laid out, which would naturally induce inputs from a variety of research paradigms. However, these inputs would need to be organised and conceptually clarified in a manner suitable for designers. This is the aim of the present thesis —the provision of a suitable descriptive framework of the reading process.

3.4 Conclusions and the way forward

It has been argued in this chapter that many of the problems inherent in electronic text design spring from the lack of a suitable description of the reading process. Cognitive psychology in the main but information science as well, has been criticised for providing unsuitable levels of abstraction at which to describe the human behaviour relevant to design. This is, however, less a criticism of either discipline but more an indictment of human factors researchers' failure to provide their own theories. Barber (1988) remarks that ergonomics as a discipline has relied so heavily on theories borrowed from other disciplines that members of the human factors community see no need to develop their own. A case study has been used to highlight the typical problems resulting from this approach and some differing views of human factors practitioners on the role of the discipline in design have been highlighted.

The practical question then is what would a human factors practitioner have added, for example, to the ADONIS design to make it more usable had he been involved as early as the specification stage? The simple truth of the matter is that deriving a more specific set of ergonomic criteria from the literature would have been difficult. The specification clearly included reference to most obvious variables. What would have been required to improve ADONIS is a user-centred design process involving iteration through prototypes and evaluations until satisfactory design targets were met.

The problem with this approach is that it is costly in terms of time and resources. What needs to be included is some means of constraining the number of iterations required. This is best achieved by ensuring that the first prototype is as close to the target as possible. Of necessity this would have involved carrying out task analyses of readers interacting with journals and searching for articles. Output from such work would have been fed back to the designers to guide decisions about how the prototype interface should be built. Subsequent evaluation would then have refined this to an even better form.

It is almost certain that such work would have led to a better design than the current one, from which we can conclude that the type of knowledge generated by task analyses and prototype evaluations is directly relevant to design. The questions then become, what form of knowledge is this, at what level is it pitched, and, more importantly, can a generalised form be derived to cover all reading situations regardless of the text and task?

In the present thesis an attempt to provide answers to questions of this type will be made. The primary means of providing them will be to examine the inputs made by the author to the design of a real-world hypertext system developed at HUSAT as part of a British Library funded project called Quartet (Tuck *et al.*, 1990). By using this system as a background it is possible to identify the type of human factors inputs needed and found to be useful in a real design project.

Given the system's form as a hypertext academic journal database, much of the work will concentrate on this text type. If the results are to be generalisable however it is important to know how this text type differs from or is similar to others. Without such knowledge it would not be possible to make any meaningful generalisations about electronic text design from any one study or series of studies on a text. Unfortunately, there is as yet no agreed classification scheme for describing the similarities and differences between texts. To overcome this, a suitable classification scheme must be developed as a first stage of the work in deriving a framework for designing electronic texts. This is in line with other views. As de Beaugrande (1981) puts it:

> To adequately explore reading , a necessary first step is a firm definition of the notion of 'text'– it is not just a series of sentences as one is often required to assume." (p. 297).

He goes onto say:

> "It follows that reading models will have to find control points in the reading process where text-type priorities can be inserted and respected" (p. 309).

To this end, the question of text type is addressed first in this thesis and an investigation into readers' own classification systems of the world of texts is reported in the following chapter. This will be used to provide a basis to subsequent work and offer a means of generalising beyond the particulars of any one particular text type.

CHAPTER 4

TOWARDS THE CLASSIFICATION OF TEXT TYPES

4.1 Introduction

The need for a useful typology of texts to aid distinctions between potentially suitable and unsuitable electronic texts has been identified recently in the domain of hypertext research (Brown, 1988; McKnight *et al.*, 1989). Such a typology would presumably provide a basis for distinguishing between the uses to which different texts are put and suggest the interface style required to support their hypertext equivalents.

At first glance it may appear that such a typology would be relatively easy to develop. Obvious distinctions can be drawn between fiction and non-fiction, technical and non-technical, serious and humorous, etc., which discriminate between texts in a relatively unambiguous manner. However, such discriminations are not necessarily informative in terms of how the text is used or the readers' views of the contents, aspects which should be apparent from any typology aiming to distinguish meaningfully between texts.

The categorisation of texts has received some attention from linguists and typographers (Waller [1987] provides an excellent review). For example, de Beaugrande (1980) defines a text type as

> "a distinctive configuration of relational dominances obtaining between or among elements of the surface text, the textual world, stored knowledge patterns and a situation of occurrence" (p. 197)

and offers the following illustrations: descriptive, narrative, argumentative, literary, poetic, scientific, didactic and conversational. However, de Beaugrande freely admits that these categories are not mutually exclusive and are not distinguishable on any one dimension. Waller adds that it is not at all clear where texts such as newspapers or advertisements fit in such a typology and proposes instead analysing text types in terms of three kinds of underlying structure:

- topic structure, the typographic effects which display information about the author's argument e.g., headings;
- artefact structure, the features determined by the physical nature of the document, e.g., page size;
- access structure, features that serve to make the document usable e.g., lists of contents.

While much of this work is interesting it is concerned less with the readers and their conceptualisation of the text than with lay-out, presentation and writing style. Inter-disciplinary boundaries are not always clear though, as cognitive psychologists have taken increasing interest in the relationship between so called 'typographical' features and the reading process (e.g., Hartley 1985) and typographers look to psychology for theoretical explanations of typographic effects. Ultimately, typographers approach the problem from the practical point of view of texts and their design, not from the perspective of better understanding the reader. Their interests may, like graphic designers lie in "getting the message across" but stop short of more fully appreciating human cognition. Thus, any typographical classification of texts is likely to differ significantly from one based purely on psychological principles.

In a more psychological vein, van Dijk and Kintsch (1983) use the term "discourse types" to describe the superstructural regularities present in real-world texts such as crime stories or psychological research reports. As noted in chapter two, their theory of discourse comprehension suggests that such types facilitate readers' predictions about the likely episodes or events in a text and thus support accurate macroproposition formations. In other words the reader can utilise this awareness of the text's typical form or contents to aid comprehension of the material. In their view, such types are the literary equivalent of scripts or frames and play an important role in their model of discourse comprehension. However, they stop short of providing a classification or typology themselves and it is not clear how this work can be extended to inform the design of electronic documents.

From a less theoretical standpoint Wright (1980) describes texts in terms of their applicative domains:

- domestic (e.g. instructions for using appliances)

- functional (e.g. work-related manuals)
- advanced literacy (e.g. magazines or novels)

She uses these categories to emphasise the range of texts that exist and to highlight the fact that reading research must become aware of this tremendous diversity. This is an important point and the impetus behind the present search for a classification scheme. Research into the presentation and reading of one text may have little or no relevance to, and may even require separate theoretical and methodological standpoints from, other texts. It is clear from the literature cited that researchers from a range of disciplines see the concept of text types as valid. The issue then is to distinguish meaningfully between texts in terms suitable for present purposes.

In order to develop a more suitable classification for electronic text design it was decided to carry out an investigation of reader's views. The study reported here marks a first attempt at developing a classification according to *reader*-perceived differences. The aim is to identify how readers describe their uses for, and of, different texts and the extent to which different readers perceive various texts in similar ways. In so doing it is hoped that any emerging classification criteria will provide clues as to how electronic documents can best be designed to suit readers, an intention not attributable to any of the aforementioned categorisations.

4.2 Distinguishing between texts: the repertory grid approach.

It was not immediately obvious how a researcher interested in text classification schemes should proceed from this point. Other researchers such as those cited above rarely make explicit the manner in which they derived their classifications. It seems that regardless of theoretical background, most, if not all, have based their schemes on their subjective interpretations of the range of texts in existence. True, their classifications often seem plausible and the knowledge and expertise of some proposers is extensive, yet it is difficult to justify such an approach in the present context.

Recognising the need for a more objective means of classification is one thing, identifying a suitable technique which enables this is another. Since the intention is to derive a reader-relevant classification this necessarily requires some means of measuring or scoring readers' views. The available options are limited to techniques such as interviewing, questionnaires or developing some form of sorting task for

readers to perform on a selection of texts. The case for using a questionnaire is flawed by the absence of any psychometrically valid questionnaire on this subject and the impossibility of developing one within the timescales of this thesis.[1] This left two main options: interviews and sorting tasks.

Interviewing individuals is a sure method of gaining large amounts of data. However making sense of the data can prove difficult both in terms of extracting sense and in overcoming subjective bias on the part of the interviewer. While the latter problem could be lessened by using two skilled interviewers rather than one, the former problem is more difficult to guard against in this context. It was felt that the nature of the interview, i.e., text classification, would be sufficiently abstract as to cause interviewees problems in clearly articulating their ideas in a manner that would support useful interpretation of the data. With these issues in mind it was decided to consider some form of sorting task.

A sorting task would involve presenting subjects with a variety of texts and asking them to describe their own particular classifications. Repertory grid analysis was eventually chosen as the most suitable technique of this type for eliciting suitable data. Developed by George Kelly (1955) as a way of identifying how individuals construe elements of their social world, Personal Construct Theory (PCT) assumes that humans are basically "scientists" who mentally "represent" the world and formulate and test hypotheses about the nature of reality. There is no need to develop a detailed account of his theory here as it has been adequately presented elsewhere (e.g., Kelly, 1955; Bannister and Fransella, 1971).

The repertory grid technique has been used for a variety of clinical and non-clinical applications (e.g., studying neurotics: Ryle, 1976; magistrates' decision making: McKnight, 1981; categorisation analysis: Coltheart and Evans, 1982; job analysis: Hassard, 1988) and has been applied to the domain of Human Computer Interaction, particularly with respect to elicitation of knowledge in the development of expert systems (Shaw and Gaines, 1987). The technique is no longer

---

[1] A questionnaire developed according to sound psychometric principles is a lengthy task involving stages of item generation, selection, piloting and analysis, possibly through several iterations before a reliable and valid tool is developed (Oppenheim 1966). This distinguishes questionnaires from the more loosely created "questions on a page"-type surveys common to human factors.

inextricably tied to Kelly's theory of personal constructs and its use as an analytic tool does not require acceptance of the model of man which Kelly proposed (Slater, 1976). However the terms Kelly used have become standard. Therefore we may describe the technique as consisting of *elements* (a set of "observations" from a universe of discourse), which are rated according to certain criteria termed *constructs*. The elements and/or the constructs may be elicited from the subject or provided by the experimenter depending on the purpose of the investigation. Traditionally both are elicited from the subject, however when a number of individuals are to be compared, it is considered best to provide the elements and let the subjects express their own constructs. Regardless of the method, the basic output is a grid in the form of n rows and m columns, which record a subject's ratings, usually on a five or seven point scale, of m elements in terms of n constructs.

The typical elicitation procedure involves presenting a subject with a subset of elements and asking her to generate a construct which would meaningfully (for her) facilitate comparison and discrimination between these elements. The aim is to elicit a bi-polar dimension which the subject utilises to comprehend the elements. A common example in the literature involves asking a subject to generate a construct about several people she knows. The subject might respond by generating a construct such as "like — dislike". Then all the elements (people) are rated according to this construct on a five-point scale where 1 might signify "strong like" and 5 might signify "strong dislike" etc. As constructs are elicited and all elements subsequently rated on these, a picture of the subject's views and interpretations of a part of her world emerges.

The advantages of the technique lie in the fact that while it focuses on the individual's subjective assessment of the world, it does so in a way that (given certain scaling assumptions) readily lends itself to statistical analysis. Where subjects generate their own constructs there is unlikely to be confusion with test concepts or the terminology employed. Researchers employing this technique can expect to obtain as rich a variety of data as may be obtained from in-depth interviews but in a form that can be more easily numerically treated. In the words of Slater (1976):

> "repertory grid technique appears to offer the flexibility and individual focus

characteristic of projective techniques, while also retaining the precision and quantifiability of standardised tests" (p. 9).

4.3 Method[2]

4.3.1 Subjects

Six subjects (five male, one female) had grids elicited. All were human factors professionals except for one, a mature student studying Ergonomics and working during the summer vacation at HUSAT.

4.3.2 Stimulus materials

Elements were identical for all subjects and were selected on the basis of likely usage by the sample, as judged by the experimenter and several colleagues on the basis of discussions and a brief informal survey of similar professionals. They consisted of nine texts:

— a newspaper (The Independent)
— a manual (MacWrite Users Guide)
— a text book (*Designing the User Interface*: Ben Shneiderman)
— a novel (*Steppenwolf*: Herman Hesse)
— a journal (Behaviour and Information Technology)
— a catalogue (Argos Catalogue Spring 1988)
— a conference proceedings (CHI '88)
— a magazine (*M* Magazine: The Observer Colour Supplement)
— a report (HUSAT Memo)

4.3.3 Procedure

Grids were elicited individually in a quiet office. Subjects were given a brief introduction to the aims of the study and the nature of the repertory grid technique.

---

[2] The repertory grid study reported here is solely the work of the present author.

Nine elements were presented to them to examine and with which to confirm their familiarity. Constructs were elicited using the minimal context form (Bannister and Mair, 1968) which involves presenting subjects with three elements, known as the triad and asking them to think of a way in which two of these are similar and thereby different from the third. The triads were presented according to a predefined sequence arranged so that no pairings of elements were repeated. Subjects were expected to produce one construct per triad, however in practice several subjects produced more than one or were unable to produce any from a given triad. In the case of more than one being elicited these were noted by the experimenter and used in turn. Failing elicitation of a suitable construct from a triad the subject was given the option of receiving another triad or generating a construct spontaneously.

When a meaningful construct was generated the two poles were written on cards and placed either side of a 1–5 rating scale on the desk. Subjects then rated all the texts according to the construct, physically placing texts at some point on this scale according to their perception of its agreement with one or other pole. Once confirmation was obtained that subjects were satisfied with this arrangement the ratings were noted and the next construct elicited. The procedure halted when subjects felt that they were unable to generate further constructs or expressed a direct wish to finish.

4.4 Results

The results were analysed using the Shaw (1980) FOCUS program. This program utilises the city block metric rather than the common Euclidean metric (see Shaw, pp. 159–160, for a discussion of the rationale) and has the advantage of retaining the original construct labels of the subjects which tend to be lost with other analysis programs as emerging factors are renamed. FOCUS involves a two-way cluster analysis that systematically reorders the columns and rows of the data matrices to produce a focused grid with minimal variation between adjacent elements and constructs.

A focused grid for one subject is presented in Figure 4.1. The grid consists of the raw ratings made by the subjects with the element list above and the construct list below. The FOCUS program automatically reorders these to give the minimum total distance between contiguous element and construct rating columns. Dendrograms

are constructed by joining elements and constructs at their appropriate matching levels.

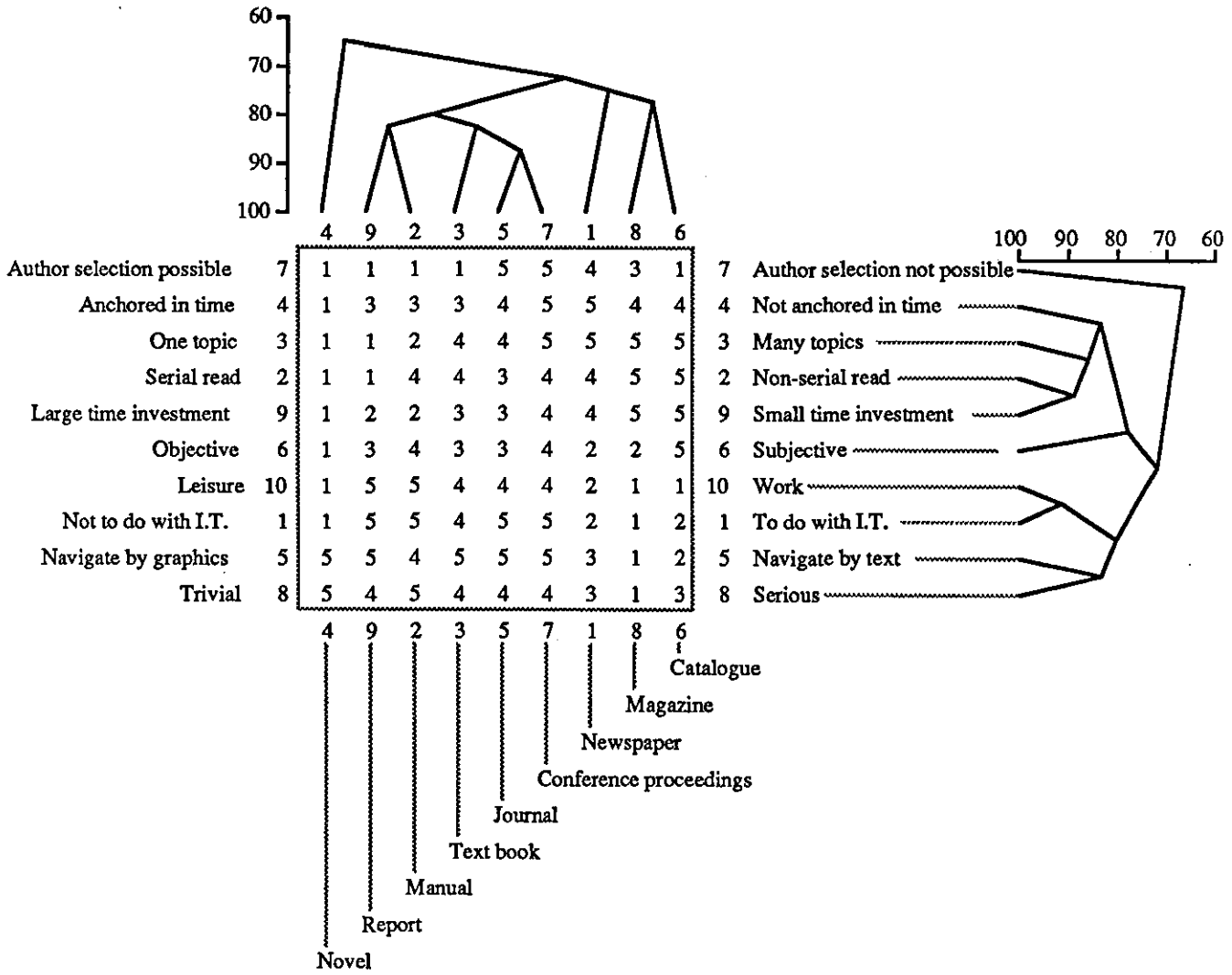


Figure 4.1 FOCUSed grid for one subject

In Figure 4.1 the element dendrogram is on top and the construct dendrogram is to the right of the reordered ratings. The matching levels for both are shown on adjacent scales. The highest match between any two of the n elements or m constructs becomes cluster $n+1$ or $m+1$ and so forth until all elements or constructs are included at cluster $2n-1$ in the case of n elements, and $2m-1$ in the case of m constructs. Therefore, as the matching level decreases the cluster number increases.

High matches indicate that the relevant elements share identical or similar ratings on the majority of constructs or the relevant constructs discriminate identically or similarly between the majority of elements. Thus in Figure 4.1 it can be observed that elements 5 (journal) and 7 (proceedings) match highly and that element 4 (novel) is least similar to the others. Constructs 10 (Work – Leisure) and 1 (To do with I.T. – Not to do with I.T.) offer the highest match among the constructs elicited, while construct 7 (Single author selection possible - not possible) is the lowest match. In other words the journal and the conference proceedings are seen as very similar to each other but very different from the novel by this reader, and every time a text is described as work related it also tends to be described as being about information technology. By proceeding in this manner it becomes possible to build up a detailed picture of how an individual construes texts.

In the present study all six grids were analysed (focused) together as one large grid. Shaw has developed SOCIOGRIDS as a means of comparing more than one grid elicited using identical elements but this is designed to give a measure of commonality between individuals by focusing every possible pair of grids and presenting a mode grid which represents the most highly matched constructs between subjects. In this way socionets can be drawn indicating how well the group cluster and which individuals, if any, are isolated from the remainder of the group. This was not an aspect which was directly relevant to the present investigation and it was therefore felt that FOCUS offered a more suitable means of analysis. Both methods are based on hierarchical cluster analysis and fine distinctions between them obscure the fact that the repertory grid's strength as a technique ultimately lies more in the skill of the person(s) interpreting the output than the particular method of analysing the data. By presenting actual dendrograms in the results it is possible to convey the interpretations of the present author. The focused grid is too large to reproduce completely here, therefore the element and construct trees are presented separately in Figures 4.2 and 4.3.

4.4.1 Elements

The elements clustered into three distinct groups (see Figure 4.2). These were the work related, the "news" type texts and the novel. The highest match was between the conference proceedings and the journal (90.2%) followed by the newspaper and

the magazine (85.1%). Basically this means that any time, for example, the journal was rated as being high or low on construct X then the conference proceedings were rated similarly. The textbook and report both joined the first cluster at more than the 82% matching level. This cluster eventually incorporated the software manual at the 62% level suggesting that while this manual shared some of the ratings of the other elements in that cluster it was noticeably different from them on certain constructs.

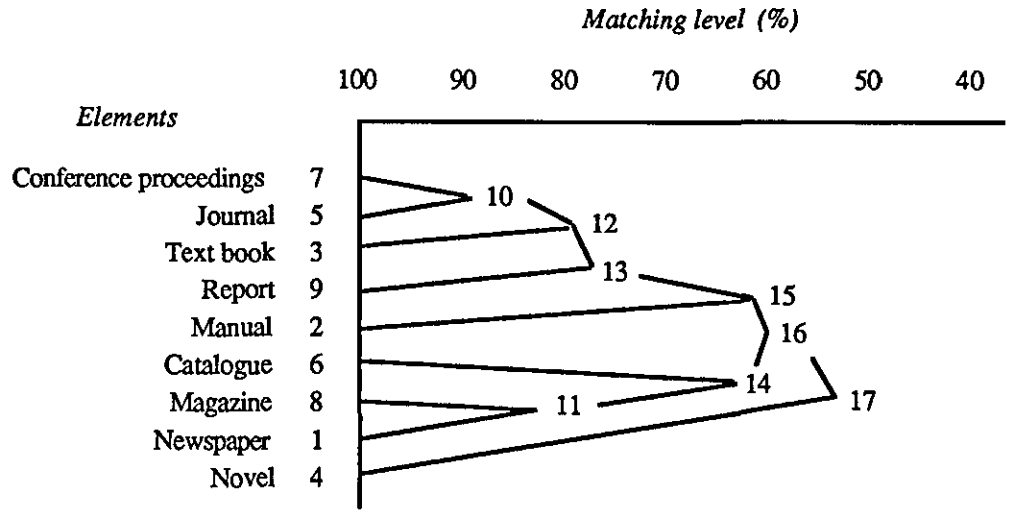


Figure 4.2 Dendrogram of element clusters for all subjects.

The catalogue matched the newspaper and magazine at 69.4% which suggests that it is perceived as similar in many ways to those types of text. The novel however was the last element to be incorporated in a cluster, only linking with other elements at the 53.2% level, by which time all the other elements had formed one large cluster. This suggests that it is a unique text type among all these elements.

4.4.2 Constructs

Fifty-four constructs were elicited from this sample. In order to ensure that only tight clusters were identified a minimum matching level of 70% was defined as the criterion. Thus any constructs that matched below this level were ignored for the purposes of further analysis. The construct dendrogram is presented in Figure 4.3. Three major construct clusters emerged from this analysis which are outlined below.

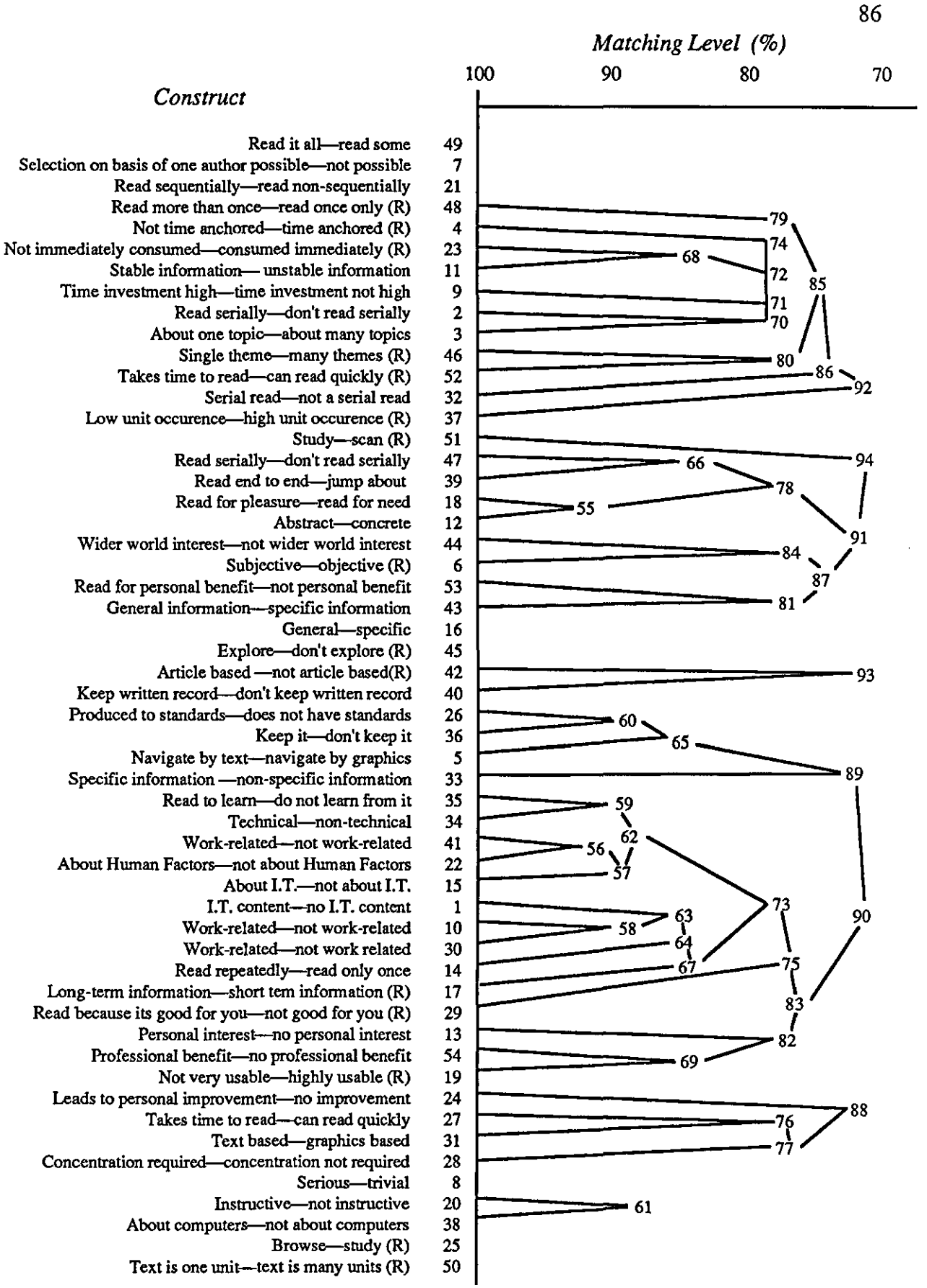


Figure 4.3 FOCUSed Contruct Set for all subjects

Cluster 1: This cluster described texts which are work-related, about human factors or Information Technology, contain technical or specific information and would be read for learning or professional purposes.

Every subject distinguished between work-related and personal reading material. All of their constructs about this distinction matched up at the 77.7% level. The highest match was between subject 3 (work-related) and subject 5 (about human factors) at 94.4%, the next highest being these with subject 2 (I.T.-related) at 88.8%, and subject 4 (work-related) and subject 1(work) at 88.8%. Obviously construing texts in terms of their subject matter and relevance to work is common to all subjects. The constructs "reading to learn" and "technical" matched at 88.8% and joined up with "work-related"/ "about human factors"/ "I.T.-related" at that level too. Also contained in this cluster were "read repeatedly" and "long-term information", matching a work-related sub-cluster at the 83.3% level. An element that was prototypical of this construct cluster was the journal. A very poor match with this cluster was observed for the newspaper and the magazine.

Cluster 2: This contained texts that were seen as personal reading material, containing general or abstract information that would be read in a serial fashion.

The highest match in this cluster was between the constructs "abstract – applied" and "reading for pleasure – reading for need" which matched at the 94.4% level. The next highest was at the 83.3% level between "serial" and "read from end to end". These pairs of constructs then joined at the 77.7% level. The constructs "personal benefit", "general information" "subjective" and "wider-world interest" all matched at the 77.7% level. These sub-clusters all joined at the 72.2% level.

These constructs suggest that certain texts are seen as more personal than work-related and contain information that is general in nature or subjectively interesting. The presence of constructs that indicate they are read in a serial fashion would indicate texts that are not intended for reference but for complete reading. A text that closely matched most of these descriptors was the novel. A very poor match with these constructs was the catalogue.

Cluster 3: This cluster described texts that were seen as having one main subject or

topic, the content of which is stable and requires a high time-investment to read. Such texts are also characterised by serial reading.

The highest match in this cluster was between the constructs "not immediately consumable" and "contains stable information" which matched at the 83.3% level. The constructs "read serially" and "one topic" matched at the 77.7% level, as did "time investment to read is high" and "single theme". The constructs "read more than once" and "not anchored in time" matched at 77.7% and all these constructs were joined by another construct "serial" at this level too. The final construct in this cluster above the criterion level was "low occurrence of separate units in the text" which joined all of the other constructs at 72.2% level. A text that closely matched these constructs was the novel; the newspaper and magazine were typically rated as the opposite of these.

Given the high cut-off point for matching constructs that was adopted, it is not surprising that several constructs remained outside the clusters. However, in all, only 8 failed to match with any other construct at this point. These were: Read it all – Read some; Select on the basis of author – Cannot select in this way; Sequential – Non-sequential; General – Specific; Explore to see if it has the answer – Know that it has the answer; Serious – Trivial; Browse – Study; Text is one unit – Text is many units. Since several of these are similar to constructs that matched highly in particular clusters their omission may appear surprising. This raises two points. Firstly, the cut-off point for matching levels is decided by the experimenter and is therefore somewhat arbitrary. A matching level of 40% could have been decided upon which would then have encompassed all constructs. The point of setting a relatively high matching level such as 70% is to extract comparatively tight clusters that share much commonality. Secondly, even though the terms "sequential" and "browse" may seem very similar to other constructs such as "serial read" or "study" the manner in which they deal with the elements is the main factor in grouping constructs. Though terms might appear similar, if they do not distinguish between elements in the same way then they are unlikely to be referring to exactly the same concept.

4.5 Discussion

A large amount of data has been reduced to a more manageable level by the FOCUS

package. The results demonstrate that people's manner of construing texts is complex and influenced by numerous factors. Clear distinctions between texts such as "fiction and non-fiction" have been shown to be simplistic and superficial. On a psychological level individuals are more likely to make distinctions in terms of the type of reading strategy that they employ with a text, its relevance to their work or the amount of information that a text contains.

The construct clusters that emerged, though reasonably tight are not clear-cut. Certainly the cluster pertaining to work-related texts that contain technical and specific information and are read in order to learn or gain professional benefit is intuitively sensible given the subject sample employed. Such a cluster highlights the easily overlooked fact that much of our reading is inextricably tied up with our work and is not just a leisure activity. However the other two clusters are not so distinctive and share many similarities. Both relate to similar reading strategies (serial or sequential) with one cluster emphasising the number of topics or themes in the text and the other the fact that certain texts are read for pleasure or personal gain. These two clusters are adjacent in the focused grid and join up eventually at the 61% matching level, suggesting general similarities between them. Had a lower cut-off point been decided upon it would have been possible to interpret the construct results in terms of two main clusters rather than three.

While the terms or descriptors employed and their similarities or differences (the face validity of the output) are interesting, it is their treatment of the elements that is ultimately important. Here the results are more specific. The elements: text book, journal, conference proceedings and report all match very highly forming a particular cluster of text types. The magazine and newspaper also match very highly. These are reasonable groupings between what may broadly be termed "work" and "leisure" texts. The novel is the one text type that matches least well with all the others and once again, this appears sensible. Examining the constructs that distinguish between these texts can shed more light on the classification criteria employed by these subjects.

The journal and text book types are described, unsurprisingly, as work-related, about human factors or I.T., containing specific or technical information and are read for professional benefit or in order to extract specific information. They are likely to be read more than once and be of long-term rather than immediate use or

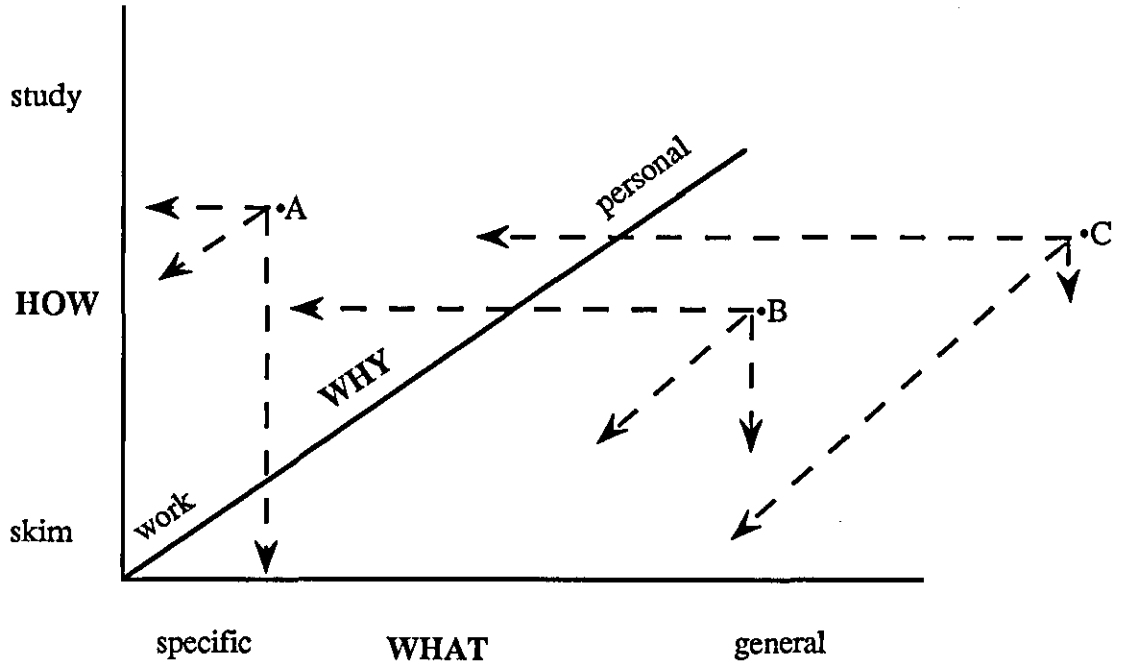
relevance. This distinguishes them from the other two element clusters which are more likely to described as read for leisure and containing general, subjective, or non-technical information. The novel is further distinguished from the newspaper and magazine by the need to invest a lot of time and read it serially and completely.

Elements therefore seem distinguishable on three levels:

- **Why** they are read e.g., for professional or personal reasons, to learn or not, out of interest or out of need, etc.
- **What** type of information they contain e.g., technical or non-technical, about Human Factors or not, general or specific, textual or graphical, etc.
- **How** they are read e.g., serially or non-serially, once or repeatedly, browsed or studied in depth, etc.

Re-examining the constructs it is possible to classify virtually all of them as referring to one of these aspects of the text. In fact the only constructs that do not seem to refer to one of these attributes are Constructs 40 (keep a written record – don't keep a written record), 36 (keep it – don't keep it) and 19 (highly usable – not very usable) which are either very specific to the individual or bear little obvious semantic resemblance to any other construct with which they may match.

By viewing text types according to the various attributes of these three levels of use it is easy to distinguish between e.g., a novel and a journal. The former is likely to be used for leisure (**Why**), contain general or non-technical information (**What**) and be read serially (**How**),whereas the latter is more likely to be used for professional reasons (**Why**), contain technical information which includes graphics (**What**) and be studied or read more than once (**How**). This approach facilitates a classificatory system as in Figure 4.4.



A: Specific information, to be studied, work-related text, e.g., a relevant scientific article

B: Less specific than A, can be read quickly, general work-related text, e.g., technical magazine

C: Non-technical, can be skim read, personal interest, e.g., newspaper.

Figure 4.4 A three-way classification of texts based on How, Why and What attributes.

Here, three texts are distinguished according to their positions relative to the **Why**, **What** and **How** axes. The descriptors study-skim, work-personal and general-specific may vary and are only intended as examples of common constructs employed by the present sample. Other samples may employ very different terms. However, these are still likely to be descriptors that pertain to the attributes Why, What and How.

Obviously there is an individualistic aspect here. The same text may be classified differently by any two readers. Thus a literary critic is likely to classify novels differently from a casual reader. Both might share similar views of **how** it is to be

read (e.g., serially or in-depth) but differ in their perceptions of **why** it is read or **what** information it contains. The critic will see the novel as related to work while the casual reader is more likely to classify it as a leisure text. What it contains will differ according to the analytic skill of the reader with a critic viewing e.g., Joyce's *Ulysses* as an attempt to undermine contemporary English and the casual reader (if such exists) seeing it as an powerful stream of consciousness modern work. Neither is wrong; in fact both are correct. Any classification of texts based on psychological criteria must, by definition, allow for such individual differences in perception.

That readers may vary their classification of texts according to tasks (i.e., within-subject differences) was also apparent from the comments of subjects in this study. Several of them remarked that some texts could possibly be classed as work-related and personal reading depending on the situation. An obvious example of this occurs when someone reads an academic article that is both relevant to one's work and intrinsically interesting in its own right. For individuals whose professional and personal interests overlap such an occurrence was common. The present categorisation of texts allows for this by placing emphasis on the motivation for reading (the **Why** axis).

Allowing for both between and within-subject variance reflects the underlying psychological complexity of the categorisation. A more rigid classification based on demonstrable objective distinctions is likely to have either a very limited sphere of application outside of which it loses relevance or be as simplistic as distinctions of the form paperback – hardback or fiction – non-fiction, etc. The present classification does not therefore provide a genuine typology of texts, but it does offer a way of distinguishing between them and analysing readers' perceptions.

How does all this relate to the development of electronic text systems? Obviously the desire for a stable classification of texts that could have their electronic equivalents empirically decided upon and standardised remains unfulfilled. In fact, this study suggests that such a typology is probably not feasible. Theoretically, at least, one could seek to determine the complete membership of each attribute set by identifying all possible Hows, Whys and Whats and subsequently plot the range of texts that match any combination of the three. However, such a level of analysis is probably too fine-grained to be worth pursuing. Even if, for example, all possible

reading styles and strategies were identified and documented (no mean feat in itself) thereby specifying the complete set of How variables, it is not at all clear how a similar approach could be usefully employed with the Why and What aspects.

The classification that is proposed here is best seen as a simple representation of the factors influencing readers' perceptions of texts. Characterising text according to How, Why and What variables provides a useful means of understanding the manner in which a given readership is likely to respond to a text. So, for example, it could be used for understanding the similarities and differences between a telephone directory:

Why: to contact a friend;
What: specific numeric data;
How: skim and locate;

and a novel:

Why: leisure;
What: large text, no graphics;
How: serial, detailed read;

and thus quickly facilitate sensible decisions about how they should be presented. Given what we know about reading from screens and HCI, the position of a novel in such a classification system would suggest that an electronic version would not be used frequently whereas an electronic telephone directory may usefully be designed to ease the process of number location or aid searching with partial information.

At a gross level, such a classification may serve to guide decisions about the feasibility of developing a usable electronic version of a text type. Where the likely readership is known this can act as a stimulus to meaningful task analysis to identify how best to design such texts. In the example of the novel above, task analysis might reveal that the novel is required for teaching purposes, where sections need to be retrieved quickly and compared linguistically with other sections or writers. Obviously this would alter the How and Why attributes of the previous classification, indicating that an electronic version is now more desirable.

The classification system can be developed by further work on increasing our understanding of each of its attributes. By knowing more about why individuals access texts, how they use them in terms of reading strategies and the distinctions they make between the information type presented it should prove possible to be more specific about the question of text types. Traditionally the How question has been the domain of the psychologist, the What question the domain of the typographer and the Why question has been largely ignored. The results from this study strongly suggest that this is a mistake. Texts exist as entities in a living psychological space where motivations, interests, anxiety etc. are known to influence the processing strategies of the human (Eysenck, 1983). To ignore such factors is to misrepresent fundamentally the reader and user of texts and impoverish the findings of any investigation.

4.6 Conclusions and the way forward

The variance in texts that readers regularly utilise was identified as a natural starting point for any analysis of the reading process and an area needing specific attention if advice is to be tailored for designers of electronic text systems. The present investigation attempted to develop a means of accurately describing texts in a manner suitable for discussing the design of electronic versions. To this end, a description based on three reader-perceived characteristics is proposed: the Why, What and How aspects.

These aspects represent readers' own classification criteria and also offer a means of describing texts in a way that is directly related to designing electronic versions, i.e., if a book is accurately described according to these criteria it should lead to specific issues for consideration in design in a way that would not be the case for a description based on more traditional criteria. Specifically, by focussing attention on how a text is read the classification immediately leads to a consideration of task issues; by examining what a text contains the questions of content and structure are addressed; and by focussing on the why aspect the context and motivation of reading are highlighted. Furthermore, by appreciating the differences in texts in these terms one is in a better position to judge the likely relevance of findings on one text to another.

Thus, as a first stage in gaining an adequate descriptive framework of the reading process as it pertains to electronic text design, the classification of texts seems to have borne fruit. The real test of this approach however, is the extent to which meaningful data can be derived from such a classification. In the following section, this is empirically tested on two text types.

CHAPTER 5

ANALYSING READING ACCORDING TO WHY, WHAT AND HOW CHARACTERISTICS

5.1 Introduction

The classification of text proposed in the last chapter is a useful starting point for a consideration of the design issues for electronic text. It is based on reader-relevant criteria and reflects cognitive and behavioural aspects of the reader-text interaction rather than any less relevant but more common criteria such as publication genre or subject matter. Focussing attention on these three aspects immediately brings forth issues related to task, motivation for reading, and readers' models of the information space – factors certain to be of importance in the ultimate success or failure of a presentation medium.

It is a simple enough matter to describe any text according to these three criteria if the description consists only of the type of one-liner provided in the examples of the last chapter. However, such descriptions are not enough to provide a firm basis for specifying software where details of a more precise nature are usually required (Easteal and Davies 1989). Furthermore, merely describing texts in this way as a result of introspection or best guesses on the part of the designer or human factors expert is far from optimum (though such uses might be appropriate for the initial consideration of issues in a design prior to formal specification). What is required therefore is the demonstration that this classification scheme can be utilised to gather evidence of reader behaviour if necessary and that the resulting output has relevance to system design. The present chapter examines these questions.

5.2 Selecting the text types to analyse

In order to provide breadth to this investigation it was decided to analyse two distinct text types. The degree of distinction was judged by reference to the previous study as well as by considering the likely tasks the selected texts would support. The final selection was also motivated by the desire to use texts that are likely to be considered for electronic versions in the real world. A constraint was the need for both text types to be relevant to a similar sample of readers. With these factors in mind, academic journals and software manuals were identified as most suitable.

The main reason for using academic articles has been alluded to already. Much of the

work in this thesis was carried out in parallel with, and as part of, a development project building an electronic journal. The present investigation therefore had direct relevance to real design issues for the team designing the application. Software manuals were selected as it was felt that they are a frequently used text for which electronic versions partly exist (in the form of on-line help facilities) and are likely to be increasingly available electronically. Furthermore they satisfy the criterion of distinction from journals as a text type according to the results of the repertory grid study thereby offering a useful test of the breadth of application for the classification scheme.

5.3 Presenting journals and manuals electronically: a brief history

The idea of presenting academic journals in electronic form is not new. Indeed such journals have been empirically investigated at HUSAT since the early 1980s. The British Library–funded BLEND project examined the potential for information technology to support an electronic journal that not only presented the text on screen but facilitated submission and refereeing of papers electronically also.

That project highlighted the need for improved screen technology and text manipulation facilities independently of the empirical literature that was only beginning to emerge at the time (Pullinger 1984, Shackel 1987). The advantages of electronic over paper journals lie in their ease of storage, rapid access and convenience of searching. Potential disadvantages stem from the inherent problems associated with reading from screens as well as issues of copyright.[1]

Software manuals accompany any off-the-shelf or bespoke application and it is a part of the folklore of human factors that they are invariably overlooked by most users (Carroll 1984 reviews some evidence of this phenomenon). On-line help facilities are becoming increasingly sophisticated and where it was once suggested dismissively that most help facilities were merely electronic versions of the paper manual, the development of query-in-depth (QID), context-specific presentations has now created a more favourable view of the concept of electronic manuals.

Electronic manuals need not just accompany software. Manuals for hardware and non-computer based machinery, either repair or operation instructions, are all potential applications for electronic text. When one considers estimates that nuclear submarines

---

[1] Copyright is a major concern in this area as publishers seek ways to control delivery and copying of material in electronic form. It will not be discussed in detail in the present work though it will be referred to where it places practical constraints on any recommendations made on the basis of studies reported here.

or spacecraft carry more weight in, or provide more space to, paper documentation than they do any other item (see e.g., Ventura, 1988), then the potential for electronic manuals is obvious.

As yet, few electronic versions of either document type have been developed sufficiently to be used on a regular or widespread basis though it is likely that of the two, manuals will emerge faster, if for no other reason than the economic interests underpinning technical manuals and their applications are more extensive than those of academic journal publishers.[2] Demonstration systems are emerging, particularly technical manuals for industrial applications, but little is currently known about their reception by users. Given the typical problems that are known to exist with electronic text however it is unlikely that rapid acceptance and use will occur.

5.4 Analysing journal and manual usage: a note on methods

Eliciting relevant information on such attributes as why and how a text is read is not easy. As mentioned in chapter two, process data of reading in particular are hard to obtain reliably. The standard approach of psychologists is to devise an experiment to answer any question. Unfortunately, where this proves difficult or impossible it is the question that is often considered ill-specified rather than the limitations of the experimental method that are exposed. To insist on an experimental approach here would leave the current problem of how and why texts are read largely intractable and support Wittgenstein's (1953) argument that in psychology:

> "the existence of the experimental method makes us think that we have the means of solving the problems that trouble us; though problem and method pass one another by" (p. 232).

For present purposes it was felt that the experimental method was wholly inappropriate but no one alternative technique offered the means to answer the questions being posed. However the questions were certainly legitimate therefore it was decided that a mixture of investigative procedures should be employed.

Interviewing a selection of relevant readers seemed the most suitable means of gathering the primary data i.e, why people read certain texts and what they typically expect the documents to contain. As mentioned previously, the advantages of

---

[2] That said, an electronic journal entitled *Postmodern Culture* has just been launched at North Carolina State University but it is too early to assess its success.

interviewing are that it facilitates the elicitation of data that are difficult to obtain from more formal methods, as well as supporting opportunistic pursuit of interesting issues. The problems and limitations of interviewing as a data elicitation method however, are well documented (e.g., Kerlinger, 1973). Common problems are the failure to structure the information gathering process properly so that certain topics are not asked of all interviewees or emphasis is placed on one topic at the expense of another. A further problem is the scoring or coding of what can be "messy" data in a reliable manner. It is generally agreed however that potential shortcomings with interview techniques can be minimised by structuring the interview (i.e., following a fixed agenda) and using an agreed scoring scheme for data. The former is usually easier to derive than the latter.

In the analyses reported here, the criteria derived from the repertory grid study provided a loose structure for eliciting and analysing the data. Thus a core set of issues to cover with every subject was identified and a standard means of categorising answers was obtained. The interview questions were then devised by discussion amongst the team of designers involved in building the journal database. This ensured that no important issues were overlooked and that the resultant data was of direct use to the design process. Furthermore, the experimenter is experienced in carrying out structured interviews for a variety of research purposes and has successfully employed the technique for other HCI-related studies (see e.g., Dillon *et al.*, 1988). In effect, all potential sources of weakness in the method were minimised.

Interviewing alone though would not sufficiently answer the question of how a text is read. As mentioned when considering that technique for the text classification work in the previous chapter, certain issues or topics are not easy to describe adequately using only verbal means. To obtain suitable information in the present context it was decided that simulated usage or task performance with concurrent verbal protocols would complement the structured interview approach. The basic idea here was to ask subjects to simulate their typical interaction with a text from the moment of first picking it up to the time of finishing with it, articulating what they were attending to with respect to the text at all times.

The basic idea behind this method is to elicit probable task behaviour from the subject without resorting to elaborate means of investigation or generating masses of low-level data of the type that would have emerged if subjects had been set formal tasks and their interactions recorded on video tape. Subjects here were asked to look at a selection of journals and to examine each one as they normally would if browsing them in the library. They were prompted to articulate what information they cue into when they

pick up a journal, how they decide if a seemingly interesting article is really worth reading and how they read articles that are selected for individual use. They repeated this simulation for several journals until a consistent pattern emerged.

The main data source in this method is the verbal protocol. Like interviewing, much has been written about the use of verbal protocols in psychological investigations. The main issue of contention is the extent to which they can be said to reflect reliably the speaker's underlying cognitive processes or are merely a reflection of what the verbaliser thinks is appropriate and/or what they think the experimenter wants to hear (see e.g., Nisbett and Wilson, 1977).

Ericsson and Simon (1984) have developed a framework for the use of verbal protocols and related it to current theories of articulation which suggest that for tasks where subjects are required to describe what they are doing or attending to in real time (concurrent verbal reporting), objections on the grounds of inaccuracy or unreliability of self-reports rarely apply. Problems of accuracy are more likely to occur during retrospective verbal reporting ("This is how I did it....") as human memory is fallible and subject to post-task rationalisation, or when reporting on how their own mental activities occurred. In other words, when humans report what they are doing or trying to do during the performance (or simulation) of a task, and are not requested to interpret their own thinking (as in introspection), there are no *a priori* grounds for doubting the validity of their comments.[3]

Verbal protocol data of this form are regularly elicited in HCI studies and have been used to good effect in analysing the influence of various interface variables on users' perceptions of, and performance with a system (see e.g., Mack *et al.*, 1983, Rasmussen, 1986). Concurrent verbal protocols were used here not to provide insight into the cognitive process of the reader but merely to provide a verbal accompaniment to the behavioural act of text manipulation and usage. As with interviewing, the present investigator is an experienced user of the verbal protocol method (see e.g., Dillon, 1987, Dillon and Sweeney, 1988).

To control for bias or potential limitations in the ability of one experimenter to capture

---

[3] Obviously subjects may lie or deliberately mislead the experimenter but this is a potential problem for all investigative methods requiring the subject to respond in a non-automatic fashion. The point here however is that subjects can reliably report what they are thinking (i.e., current contents of working memory) but are less likely to be able to do so for how they came to be thinking of it (i.e., what cognitive processes brought these contents to working memory).

all relevant data another experimenter was employed for the first study thus providing a co-rater for the elicited data and ensuring that all data were captured. All interpretations and conclusions were checked with this experimenter before final agreement was reached. After the first investigation, due to the high level of agreement between experimenters and the lack of difficulty in one experimenter recording all the data, this was felt to be unnecessary for further work.[4]

5.5. Analysing journals according to Why, What and How attributes.

5.5.1 Subjects

15 subjects (seven male/eight female; age range 22-34) participated in this study. All were self-described regular journal readers involved in research professionally. Though all working in the same domain (human factors) they came from one of three first-degree backgrounds: psychology, ergonomics or computer science.

5.5.2 Stimulus materials

Subjects were presented with a selection of academic journals according to their expressed list of normal reading material. The general pool of texts consisted of multiple copies of:

– Behaviour and Information Technology
– International Journal of Man-Machine Studies
– Human Factors
– Ergonomics
– British Journal of Psychology
– Design Studies
– Work and Stress
– The Computer Journal

5.5.3 Design and procedure

The basic procedure involved an interview to collect information on why subjects used journals and what types of information they thought such a text contained. Typical prompts at this stage involved variations of the Why and What aspects, persuing points as they developed and concentrating on those areas that have any impact on usage.

---

[4] The author gratefully acknowledges the contribution of John Richardson here. However the analysis of this study and the complete manual usage study reported later in this chapter are solely the work of the present author.

Subjects then interacted with a sample of relevant texts, simulating their typical usage according to their expressed reasons, articulating what they were attending to as they did so. They were prompted as necessary by the interviewer.

After describing and simulating their typical usage of the texts the interviewer described his impression of their style and sought feedback from the subject that it concurred with the subjects' own views. When agreement had been reached, i.e., the subject and interviewer both agreed that the representation of usage was accurate and adequate, the interview ended.

5.6 Results

The experimenter grouped reponses according to meaning and frequency and the results will be broadly presented in terms of the three text usage criteria: why, what and how. Other aspects that emerged as a result of the interviews and observations are included where relevant even though they may not strictly conform to these aspects.

5.6.1 Why read journals?

The most frequently stated reasons for accessing journals are summarised in Table 5.1. Virtually all subjects distinguished between problem-driven journal usage where work demands require literature reviews or rapid familiarisation with a new area, and personal usage where journals are browsed in order to keep up with latest developments in one's area of expertise or interest, the former being cited more frequently than the latter. Obviously work demands vary and periods of heavy use are matched by times of little requirement for articles.

Why use a journal?	No. of Ss.	(%)
Background material for work purposes	11	(73)
Updating one's knowledge	7	(46)
Personal interest	3	(20)
On recommendation	2	(13)
Following up references	2	(13)

Table 5.1 Stated reasons for using journals

In more specific terms, journals are accessed for: personal interest, to answer a particular question (e.g., "what statistics did the authors use and why?"), to keep up

with developments in an area, to read an author's work and to gain advice on a research problem. In other words, there are numerous varied reasons for accessing material in journals apart from just wanting to "study the literature".

5.6.2 What type of information is in a journal?

The general consensus among the present sample was that journals are a predominantly textual rather than graphical form of documentation (mentioned by 60%) which conform to a rigid style of presentation and composition (80%). Information tends to be relatively technical such that only readers versed in the subject matter could profitably read it (46%). Furthermore, presentation style is highly formalised, i.e., written in a manner unique to journals that differs from conventional prose.

Five subjects (33%) remarked that the formal style of journals was off-putting. The unnecessary use of references, the wariness of authors to express opinions and the over-reliance on statistical details in experimental reports were all cited critically and this is reflected in the actual reading styles of several subjects for this type of material (see below). Graphics in articles (e.g. tables, figures etc.) were generally viewed positively. Seven subjects (46%) explicitly stated a dislike for articles that consisted of pages of straight text. Heavily mathematical content was viewed negatively by the readers in this sample.

The concept of structure in articles was discussed with all subjects and it was apparent from their responses that most readers feel that this text type is organised around a relatively standard framework which the majority of the present respondents viewed as being of the form:

- Introduction
- Method
- Results
- Discussion/Conclusion

or of the form:

- Introduction
- Elaboration and criticism of issues
- Alternative views proposed by author
- Discussion/Conclusion

depending on whether it is an experimental or review type paper.

This order was seen as useful for reading purposes as it facilitated identification of relevant sections and allowed rapid decision making on the suitability of the article to a reader's needs. For example, poor sectioning, large method and results sections, small discussions and large size in terms of number of pages were all cited as factors that would influence a reader's decision on whether or not to reject an article.

The issue of article size is interesting. Large articles obviously require a significant time-investment which is often seen as a disincentive. Perceptions of what constituted a large or small article varied. Large articles were described as being anything from 6 to more than 30 pages long, medium length articles as being between 5 and 20 pages long and small articles being between 3 and 20 pages long. In other words what one individual rates as large, another may rate as small. Median responses suggest that articles more than 20 pages long are large and those articles that are about 5 pages long are small. Approximately 10 pages is considered to be medium length.

5.6.3 How are journals read?

All subjects simulated and described their use of journal articles. Though no two subjects provided identical descriptions there was a high degree of commonality between all subjects. As mentioned previously, all subjects confirmed the experimenters interpretation of their usage style before ending the interview. Figure 5.1 represents a generic description of usage patterns and though not totally representing any one subject, it contains no actions that all subjects did not manifest or articulate.

First, all subjects skim read the table of contents of the issue. A preference was expressed for contents printed on the front or back page which made location of relevant articles possible without opening the journal. If the reader fails to identify anything of interest at this point the journal is put aside and, depending on the circumstances, further journals may be accessed and their contents viewed as above. When an article of interest is identified then the reader opens the journal at the start of the relevant paper. The abstract is usually attended to and a decision made about the suitability of the article for the reader's purposes.

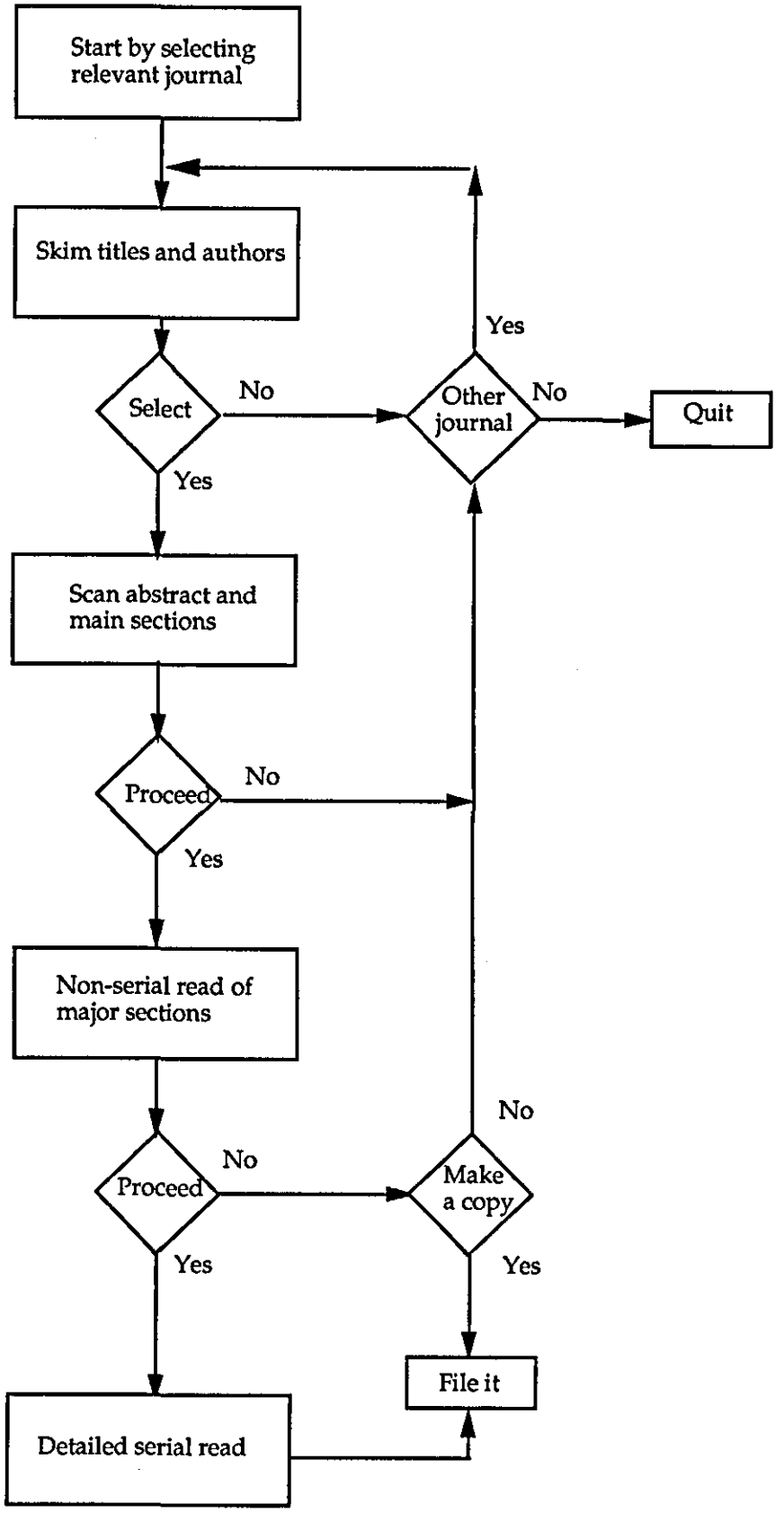


Figure 5.1 Generic model of journal use

At this point most subjects reported also browsing the start of the introduction before flicking through the article to get a better impression of the contents. Here subjects reported attending to the section headings, the diagrams and tables, noting both the level of mathematical content and the length of the article. Browsing the conclusions also seems to be a common method of extracting central ideas from the article and deciding on its worth.

By this time readers seem to have completed one cycle of interaction with the article and make a decision whether or not to proceed with it. A number of factors may lead to the reader rejecting the article. The main reason is obviously content. The reader by now has a strong impression of the type of material contained in the paper and will be able to make an informed decision on the relevance of it to his needs. How accurate this impression is remains an empirical question. If the article is heavily mathematical it tends to be rejected by the readers in this sample.

If the article is accepted (or photocopied) for reading it is likely to be subjected to two types of reading strategy. The majority of subjects (10) scan read the article rapidly and in a non-serial fashion to extract relevant information. This will involve reading some sections fully and only skimming or even skipping other sections. Typically the method and results sections of experimental papers are skim read while the introduction or introductory sections and the discussion/conclusions are read fully. Readers may highlight points or make notes at this stage.

The second reading strategy is a serial detailed read from start to finish. This was seen as "studying" the article's contents and though not carried out for each article that is selected, 11 subjects reported that they usually read selected articles at this level of detail eventually. Three subjects expressed a preference for this reading strategy from the outset over scanning though acknowledging it to be less than optimum.

While individual preferences for either strategy were reported most readers seem to use both strategies depending on the task or purpose for reading the article, time available and the content of the article. Original and interesting work is more likely to be read fully than dull or routine papers. Reading to keep up with the literature requires less "studying" of articles than attempting to understand a new area. If reading the article with a view to citing it in one of their own papers, subjects expressed a stronger tendency to read the article fully. However, even when reading at this level of detail some subjects still reported skimming particular sections that were not intrinsically relevant to their particular needs at that time.

5.6.4 Summary

Academic journals may be broadly classified according to the three criteria as follows:

> **Why**: For work reasons such as keeping up with the literature, as a source of reference and as a source of learning. They are also read for personal reasons when accessed out of interest with no immediate work requirements.
>
> **What**: Technical information about a specific domain; may have graphical components but are predominantly textual and tend to conform to a relatively standard structure.
>
> **How**: Three levels of reading:
>
> (i) quick scan of abstract and major headings;
> (ii) non-serial scan of major sections;
> (iii) full serial read of the text.

From a human factors perspective the variations in how articles are read suggest important distinctions in the type of presentation required to support the reader. Certainly, at what may be termed level three, the detailed serial reading stage, electronic presentation is likely to prove totally unsuitable with current technology. Subjects in this sample stated unequivocally that they prefer to obtain a paper copy of the text to read where and when they like, to write on and to store in their personal files for later reference. At the other two levels though, there are several interesting aspects of journal usage that have relevant design implications. The following section will outline these.

5.7 Design implications for electronic journals

At the first level all subjects attend to the Contents page of journals and prefer these to be easily accessible. It would seem therefore that a facility to scan lists of titles and authors would be desirable. These should probably be grouped as they are on paper i.e., in "issues", but the ability to scan continually should be available.

Since the full contents of the paper are not attended to at this point it is better that users are given brief information about it and offered the chance of jumping around to various sections of the text. The default mode of article presentation should not be the same as the paper equivalent. A likely presentation style based on the present findings might be: the title of the paper, the author(s), the abstract, a list of section headings that are

selectable and the references cited. Further information about the size of the article might also be useful.

Rapid browsing facilities are vital. At this initial stage of article selection and the second level (non-serial read stage) fast page-turning is common as readers jump back and forth through the article. The electronic version must support this activity by allowing both scrolling of the next page/previous page variety and rapid jumping to particular sections e.g., from the introduction to the method or the discussion. It might be desirable to facilitate jumping to "landmarks" in the text such as tables or figures too, possibly with the use of a graphical browser.

The ability to print the article at any point would be desirable as obtaining hardcopies of selected articles is a major concern of most journal readers. Keeping a record of interesting articles which can be batch printed at the end of the interaction may also be desirable. Given the observed reading styles of the present sample it might be useful to offer the facility to print sections rather than the full article. For example, readers might choose to print the introduction and discussion sections only. This would have the advantage of reducing costs of obtaining hardcopies and save on unnecessary use of paper.

Obviously these are relatively general considerations. More specific instances of journal usage e.g. searching the text for a reference or sentence about a particular point may require particular facilities and display characteristics which are not suitable for other tasks such as browsing the major sections. These are empirical issues which require further detailed research. The present discussion is not aimed at answering such questions but demonstrating how they fit into the broader perspective of text usage and the type of work that needs to be done to resolve such issues.

As an exercise in data collection the foregoing procedure can be judged as useful. The following section applies this method to a second text type to examine its suitability for other document forms.

5.8 Describing manuals in terms of Why, What and How attributes.

5.8.1 Subjects

Fifteen subjects participated in this study (six male/nine female, age range 22-41). All subjects were casual users of software manuals and drawn from the same population of

users as the subjects in the previous study.

5.8.2 Stimulus materials

A selection of five software manuals relevant to the present sample was used (as determined by an informal survey). Subjects were allowed to comment on and simulate usage of any or all of the following:

– MacWrite user's guide
– MULTICS e-mail guide
– StatWorks manual
– HyperCard user's guide
– Mac Utilities manual

5.8.3 Design and Procedure

This was identical to the design and procedure of the journal usage study. The second experimenter was not used here however as the previous study had showed that the method was manageable by one experimenter only and no obvious rater unreliability could be identified in the data.

5.9 Results

As before, the results will be broadly presented in terms of the three text usage criteria: why, what and how.

5.9.1 Why use manuals?

Subjects stated numerous reasons for using manuals though there was a large degree of consistency in their responses. These were categorised according to meaning by the experimenter and their relative frequencies summed. The categorised comments are presented in Table 4.2.

Why use a manual?	No. of Ss.	(%)
For reference/How do I do this?	11	(73)
How to get started	10	(67)
When in trouble	8	(50)
For a summary of package's facilities	5	(33)
Aid for exploring software	2	(13)
As a guide before buying	1	(7)
For detailed technical info	1	(7)

Table 5.2 Stated reasons for using software manuals

Clearly, readers have a limited range of motivations for using manuals. The three main reasons (reference, introduction and when in trouble) were all offered by at least half the subjects. These highlight the problem-driven nature of manual usage. In fact, all subjects remarked that manuals were only ever used in work or "task" domains. It should also be noted that while using a manual to get started was one of the most common motivations for use, six subjects stated that they would hate to rely solely on a manual to learn a package, preferring to use it only when absolutely necessary. Furthermore, all but two of the subjects stated that they would much rather ask someone for information than access a manual.

5.9.2 What type of information is in manuals?

The responses to questions of this nature displayed a high degree of commonality across all subjects. Invariably material was described as "technical", "specific" and "detailed". While it might be argued that the very nature of manuals is that they contain such information most subjects seemed to find this off-putting. A third of the subjects remarked that manuals were heavily loaded with jargon and information on simple actions was often difficult to locate or extract as a result.

All subjects remarked that manuals were too textual and that more graphics would often aid the user's location and comprehension of information. However, it was repeatedly pointed out that graphics should be relevant and one manual (for the MacWrite package) was much criticised for using superfluous pictures of desktops, scissors and documents.

The need for different versions of manuals which are structured according to the users' needs was suggested by four subjects. Typically it was suggested that these should consist of a manual for a "total novice" which explains how to perform very basic

procedures and a more detailed version for users who have acquired a greater degree of competence.

The structure of manuals was discussed with all subjects and responses varied between those who are aware of text structure as it pertains to this text type and those who felt it existed but had difficulty articulating their perceptions of it. Primarily, a sense of order seems to be lacking in manuals though the majority (60%) of subjects felt that there might be a progression from easy to hard as a reader moves from the beginning to the end of the text i.e., the more complex operations are dealt with towards the back of the manual. One subject remarked that while it might appear that an easy-to-hard progression exists, a structure based around command frequency was probably more frequent, i.e., commonly used commands or actions were more likely to be located at the front of the manual and less common ones at the back. Another suggested order, general-to-specific was made by two subjects. Two subjects argued that if any order such as easy-to-hard could be observed it probably existed at the task rather than the global level i.e., within sections rather than across the manual.

The perceived modal structure for manuals that emerges from the comments of the present sample is:

- contents,
- getting started,
- simple tasks,
- more complex tasks,
- index.

As this structure indicates, heavy emphasis was placed on the task as a structural unit for organising manuals. There were variations on this modal structure. For example, two subjects placed training exercises at various points in the structure, the gradation between basic and more complex tasks was extended in two cases to include an intermediate level, while others mentioned a glossary of commands, technical specifications and lists of error messages as further typical units of a manual.

Many of the problems users of manuals seem to experience are related to the question of structure. Invariably this was criticised as "poor" or "disorganised". The present sample seemed divided between those who felt that overall order was less important than the procedural order at the task level and those who were content with procedural ordering but felt that high-level ordering was unsatisfactory in many manuals.

5.9.3 How are manuals used?

The procedure for extracting information from a manual was assessed using all samples provided in order to ensure a balanced extraction of readers' behaviour. Unsurprisingly, a relatively stable behavioural pattern was observed between and within subjects over the range of manuals. Figure 5.2 represents usage styles in flowchart form. As before this is a generalisation of subjects' behaviour.

The first thing readers do is get a feel for the document's contents. Thus readers initially open the text at the contents or index sections. The majority (60%) stated that the Contents page is usually checked first and if that did not suggest where to go, the index was examined. However, it seems that much depends on the nature of the problem. If decoding an error message or seeking information on a particular command then the index is likely to provide this and will therefore be accessed first.

If more general information is sought then the contents offer a better chance of location. This highlights the extent to which book conventions have become internalised in the minds of contemporary readers. Furthermore, the index or contents list is not read in a simple pattern-matching fashion for a word template; rather the reader tries to get an impression of context from the contents i.e., what items precede and proceed it, or may have to think of other terms that might yield satisfactory information if the term being sought is not present (a common problem with technical jargon).

If the reader fails to locate anything in the contents or index that appears relevant he may either dip into the text and move about looking for relevant information or give up. The latter option appears common according to the data and represents a failure in document design that must be overcome.

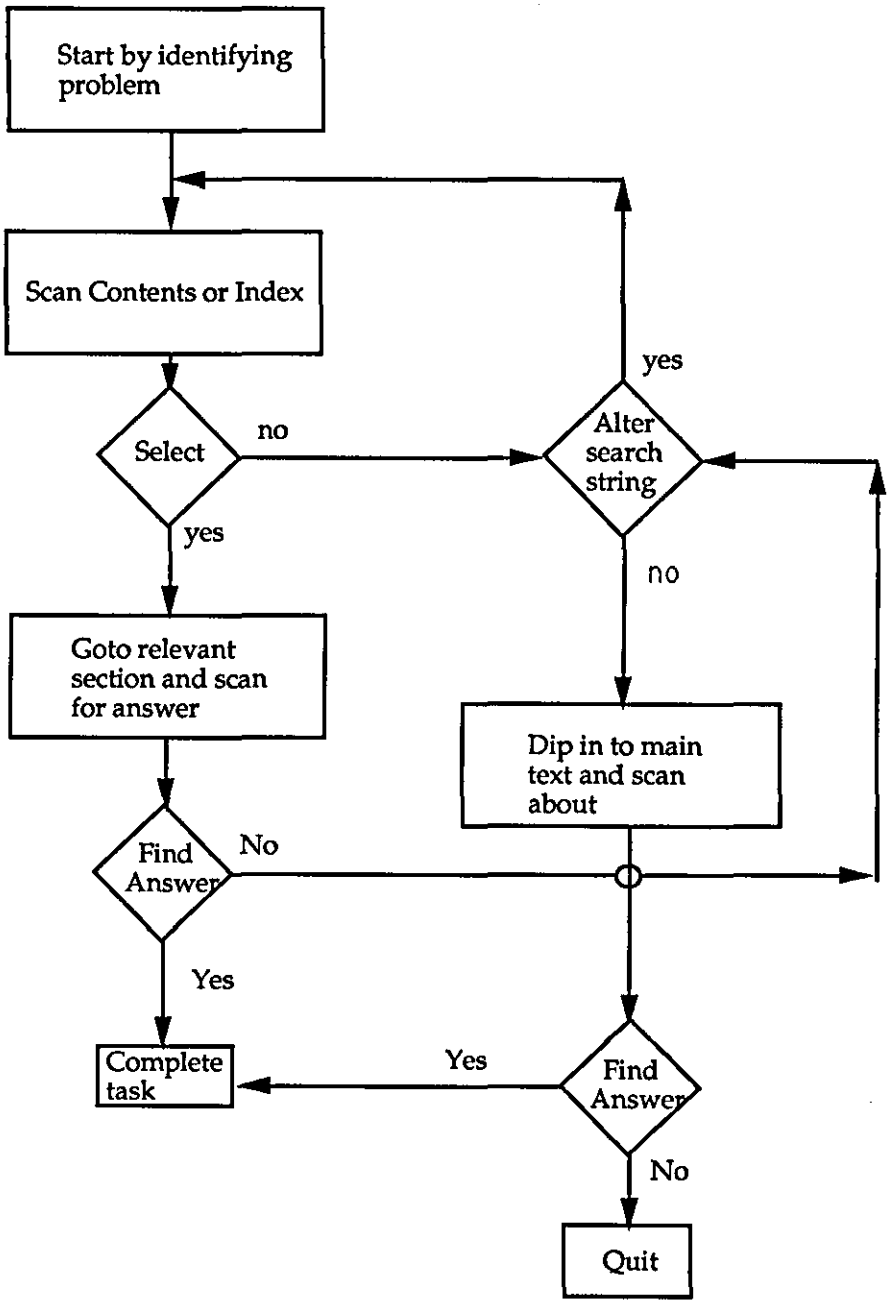


Figure 5.2 Generic model of manual usage

If on the other hand a relevant item is located then the reader turns to the relevant page and if it is not immediately obvious, will scan about looking for a relevant diagram, word, phrase etc. to indicate that the answer is there.

At this stage the reading strategy adopted by the reader depends on the particulars of the

task. If it is a simple matter of decoding an error message or finding the correct syntax for specifying command parameters then rapid scanning is adopted. If on the other hand the reader wants to perform a new sequence of actions more procedural reading will occur. However, even though the latter form will require more serial reading of the text, few subjects reported actually reading complete sections. The tendency to "get-on-with-it" seems firmly established in users of manuals and the present sample report moving freely from manual to system in order to achieve their goal.

Only three subjects manifested any tendency to read around an area or fully read a section before moving on and even these admitted that they would be tempted to skim, tend to get bored if they felt that they were nor resolving their problems and only read complete sections if all else had failed.

5.9.4 Summary

Software manuals may be broadly classified according to the three criteria as follows:

> **Why:** For task specific reasons such as troubleshooting, getting started, and for reference. Can occasionally be used for exploring software and identifying facilities or obtaining detailed technical information.
>
> **What:** Technical information, of a specific and detailed nature, often laden with jargon. Can be a mixture of text and graphics. Structure is based around task units.
>
> **How:** Problem driven. Broadly it involves checking the index or contents sections to find something relevant then dipping into and scanning sections of text. Lengthy serial reading is rare.

From a human factors perspective these are interesting data and the following section presents a brief discussion of their implications for electronic versions.

5.10 Design implications for electronic manuals

What can we say about designing an electronic software manual on the basis of these data? Potentially, the observed interrogative reading style seems highly suited to electronic presentations. However, an electronic text that merely replicated the paper manual would appear to be relatively useless. It would share all of the disadvantages of

the paper version but none of the advantages of manipulation, portability, familiarity and image quality. Since usage is so goal-oriented, large sections of a manual's contents are irrelevant for much of the time and providing a single multi-page text for the reader to search seems a less than optimum method of presenting information.

Therefore in order to ensure usability one would need to consider an alternative structure and means of access. Given the typical usage style noted above it is likely that searching facilities that supported location of information on a term or concept would be useful. It is conceivable that a thesaurus of terms would enable better searching. Detailed contents/indices with selectable items would be another enabling mechanism. None of these facilities should be beyond the scope of a well-designed hypertext application.

A continually emphasised attribute of hypertext is the ability of authors to "hide" layers of information that can be accessed by links. Thus users could conceivably follow a trail of information about a concept to whatever level of detail is required. For the curious user this could be easily achieved by actioning links. For the user who has no desire for such in-depth coverage such a presentation format could "hide" copious amounts of unwanted material. Thus we can see the emergence through templates of "versions" of manuals for different user types requested earlier. Such a structure would support certain aspects of the typical model readers possess of this information type.

Obviously this discussion of potential hypertext attributes to support electronic manuals is superficial. No mention has been made of the potential for such attributes as "interactive figures" (where users can point at sections of a diagram to gain further information on that part), scrollable pop-up windows or links to alternative procedures and methods for accomplishing a goal. Intelligent searching facilities that provided links to associated concepts when a specific term failed to elicit a satisfactory response would obviously be useful but is beyond the scope of consideration for the present thesis.

What this study has shown however is that the classification criteria can also be applied to another text type to support the elicitation of specific information on usage. Given the distinctions made between journals and manuals in the repertory grid analysis of chapter four, there is reason to believe that this form of analysis could usefully be applied to a wide range of text types.

5.11 General conclusions and the way forward.

After reviewing the literature on reading from screens it was stated that providing designers of electronic text with guidance on the basis of experimental evidence was not straightforward. The lack of a suitable description of the reading process was highlighted as a major problem in this venture. It was recognised that some form of text classification would be a useful starting point for such a description and the analysis of reader's classifications in chapter four provided three criteria for distinguishing texts. In the present chapter these criteria have been applied to two texts to examine the extent to which distinct texts are used and viewed in different ways by readers and the likely impact of such variables on electronic text design.

Though probably a non-optimum means of gathering data, the simulation based procedure used here has demonstrated that it is possible to gain reliable and design-relevant information on the reading process without recourse to sophisticated and/or intrusive tools. Subjects had few difficulties articulating how they used the relevant texts and responding to question on the What and Why aspects of reading. There was a high degree of consistency between subjects which enabled general conclusions to be drawn, yet the data clearly distinguish between text types in the characterisation of usage they suggest.

Data of this type are useful for gaining insight into the relevant design issues for an electronic document, particularly early in the product life cycle (a time when human factors inputs are typically weak) but there are limitations. It is applicable in the main, only for existing texts. For innovative information types that will surely emerge with the advent of hypermedia such a classification of task relevant issues is not possible (although subsets of the hypermedia information might be amenable to such analysis). Furthermore it is a relatively informal procedure, reliant for its value ultimately on the abilities of the practitioner more than one might like, particularly for mapping responses to interface recommendations.

However, this form of data capture and analysis is not intended to be anything more than an accurate means of initially conceptualising the issues, i.e., identifying the text type and its associated usage characteristics. By providing designers with some reliable estimates of the three aspects of usage - the why, what and how of reading - it supports reasoned constraining of the design options under consideration. More formal analyses can then be appropriately targetted at specific design issues.

In the present cases, the data only provided generic recommendations on how an electronic journal or manual should be designed to support usage. However they suggest what form prototype versions might take and what issues need to be considered further. Indeed, the journal usage data reported here proved to be directly relevant to a hypertext journal database design at HUSAT and also highlighted the issue of structure as important to its development. This resulted in attention focussing on readers' models of articles and led to two experiments which influenced the design of the database. This work is covered in detail in the next chapter.

CHAPTER 6

STRUCTURE IN TEXTS: READERS' MODELS OF INFORMATION SPACES

6.1 Introduction

An issue that emerged as important in both the journal and manual usage studies of chapter five is the notion of the perceived structure in documents. As stated in chapter two, it is now possible to embody alternative structures for electronic texts that could not be feasibly supported in the relatively standard format of paper. Typically, advocates of the "new structures" approach dismiss paper as a limiting medium, claiming it demands a linear format for presentation and consumption, contrasting this with the supposedly liberating characteristics of hypertext (see e.g., Beeman *et al.*, 1987).

It is debatable whether this is a fair representation of paper (or hypertext for that matter). Though paper texts may be said to have a physical linear format there is little evidence to suggest that readers are constrained by this or only read such texts in a straightforward start-to-finish manner (Charney, 1987). For example, the journal usage study (chapter 5) identified three reading strategies in readers of academic journals, only one of which could be described as linear, and one only has to think of one's own interaction with a newspaper to demolish arguments of constrained linear access.

Since structure is obviously an issue that needs to be considered in the development of such systems it must also be accounted for in any descriptive framework of the reading process aimed at aiding designers. The present chapter therefore examines the concept of text structure as it pertains to document use more closely and reports two studies carried out by the author to inform the design of the hypertext journal database.

6.2 The concept of structure in documents

Regardless of the putative constraints of paper texts or browser-friendly attributes of hypertexts, it seems certain that readers possess some form of mental representation for a document type that provides information on the probable structure and organisation of key elements within it. In chapter two an example was given of the schemata readers seem to possess of books i.e., when we pick one up we immediately have an idea of its subject matter, size, type of contents etc. and the studies in the previous chapter confirmed this for two distinct text types.

The same might be said of a newspaper. Typically we might expect a section on the previous day's political news at home, foreign coverage, market developments and so forth. News of sport will be grouped together in a distinct section and there will also be a section covering that evening's television and radio schedules. If this can be said to hold true for all established text forms, then developers of hypertext systems need to consider carefully their designs in terms of whether they support or violate such assumptions.

Unfortunately, the term 'structure' is used in at least three distinct ways by different researchers and writers in this field. Conklin (1987) talks of structure being imposed on what is browsed by the reader, i.e., the reader builds a structure to gain knowledge from the document. Trigg and Suchman (1989) refer to structure as a representation of convention, i.e., it occurs in a text form according to the expected rules a writer follows during document production. Hammond and Allinson (1989) offer a third perspective, that of the structure as a conveyer of context. For them, there is a naturally occurring structure to any subject matter that holds together the "raw data" of that domain.

In reality, there is a common theme to all these uses. They are not distinct concepts sharing the same name but different affects or manifestations of the same concept. The main role of structure seems to differ according to the perspective from which it is being discussed: the writer's or the reader's, and the particular part of the reading/writing task being considered. Thus the structure of a document can be a convention to both the writer, so that he conforms to expectations of format, and to the reader, so he knows what to expect. It can be a conveyer of context mainly to the reader so he can infer from, and elaborate on, the information provided, but it might be employed by a skilled writer with the intention of provoking a particular response in the reader. Finally, it can be a means of mentally representing the contents to both the reader so he grasps the organisation of the text and to the author so that he can appropriately order his delivery.

It can be seen from the comments of subjects (as readers) in the journal and manual usage studies that structure is a concept for which the meanings described above seem to apply with varying degrees of relevance. Certainly the notion of structure as convention seems to be perceived by readers of journal articles, while the idea of structure supporting contextual inference seems pertinent to users of software manuals. Beyond these manifestations, research in the domain of linguistics and discourse comprehension lends strong support to the concept of structure as a basic component in the reader's mental representation of a text.

The theory of discourse comprehension proposed by van Dijk and Kintsch (1983) places great emphasis on text structure. According to their theory, reading involves the analysis of

propositions in a text and the subsequent formation of a macropropositional hierarchy (i.e., an organised set of global or thematic units about the text). From this perspective, increased experience with texts leads to the acquisition of knowledge about macrostructural regularities which van Dijk and Kintsch term 'superstructures' (see chapter 2, Section 2.9.1) that facilitate comprehension of material by allowing readers to predict the likely ordering and grouping of constituent elements of a body of text in advance of reading it.[1]

They have applied this theory to several text types. For example, with respect to newspaper articles they describe a schema consisting of headlines and leads (which together provide a summary), major event categories each of which is placed within a context (actual or historical), and consequences. Depending on the type of newspaper (e.g., weekly as opposed to daily, tabloid as opposed to quality etc.) one might expect elaborated commentaries and evaluations. Experiments by Kintsch and Yarborough (1982) showed that articles written in a way that adhered to this schema resulted in better grasp of the main ideas and subject matter (as assessed by written question answering) than ones which were re-organised to make them less schema conforming.

Interestingly, when given a cloze test[2] of the articles no significant difference was observed. The authors explain this finding by suggesting that schematic structures are not particularly relevant as far as the ability to remember specific details such as words is concerned (i.e., the ability which is measured by a cloze test) but have major importance at the level of comprehension. In their terms, word processing and recall is handled at the microstructural level, text specific organisation at the macrostructural level and general organisation of the text type at the superstructural level.

The van Dijk and Kintsch theory has been the subject of criticism from some cognitive scientists. Johnson-Laird (1983) for example takes exception to the idea of any propositional analysis providing the reader with both the basic meaning of the words in the text and the significance of its full contents. For him, at least two types of representational

---

[1] In reality the idea of superstructure appears to be more of a spin-off than a central tenet of van Dijk and Kintsch's theory. They seem to postulate three general levels of text unit: microstructures, macrostructures and superstructures but prefer to concentrate on the first two, at this time having developed their ideas on these to a greater extent than they have on the third. However, experimental work seems to confirm the relevance of the third level of structure even if its exact relationship to their comprehension theory is not precisely specified yet. As van Dijk (1980) put it when describing superstructures: "It cannot possibly be the aim of this chapter to provide a theory of superstructures. A separate book would be needed..and we would even need separate studies for the different main kinds of superstructures" (p. 109). As yet, that book is unwritten.

[2] A cloze test is a traditional comprehension test for readers that requires them to fill in the blanks within sentences taken from the text they have just read.

format are required to do this and he provides evidence from studies of people's recall of text passages that it is not enough to read a text correctly (i.e., perform an accurate propositional analysis) to appreciate the significance of that material. He proposes what he terms mental models as a further level of representation that facilitates such understanding. Subsequent work by Garnham (1987) lends further support to the insufficiency-of-propositions argument in comprehension of text.

The differences between Johnson-Laird and van Dijk are mainly a reflection of the differences between the psychologist's and the linguist's views of how people comprehend discourse. From the perspective of the human factors practitioner it is not clear that either theory of representation format is likely to lead to distinct (i.e., unique) predictions about electronic text. Both propose that some form of structural representation occurs— it is just the underlying cognitive form of this representation that is debated. The similarity of their views from the human factors perspective is conveyed in this quote from Johnson-Laird where he states that mental models:

> "appear to be equally plausible candidates for representing the large-scale structure of discourse – the skeletal framework of events that corresponds to the 'plot of the narrative', the 'argument' of a non-fiction work and so on. Kintsch and van Dijk's proposal that there are macrorules for constructing high-level representations could apply *mutatis mutandis* to mental models" (p. 381).

In other words, the issue is not if, or even how, readers acquire a structural representation of texts they read (these are accepted as givens) but what form such structures take: propositions or mental models?[3] This is not an issue of direct concern to the designer of electronic texts, what is of importance is the provision and support of document structures that aid accurate structural representations (of whatever form) in the reader's mind.

In summary then, readers claim to observe structures in documents that facilitate predictions about their likely contents and organisation. Electronic text researchers conceptualise structure as both convention and context provider of relevance to both the reader and the author. Theories of discourse comprehension support these views in general

---

[3] van Dijk and Kintsch (1983) addressed some of the Johnson-Laird criticisms by incorporating a "situation model" of the text into their theory. This is a mental representation of the significance of the text in terms of its subject matter and the central figures/elements under discussion which facilitates the application of contextual knowledge stored in long-term memory (a major weakness of their original proposition based theory). However, they still retain the three levels of text structure (micro, macro and super) as the basic elements of their theory of comprehension, giving the situation model a relatively minor role.

but can differ from each other in terms of how they account for the representation of that information in the cognitive system. For the human factors practitioner, such fine distinctions of representational form are unlikely to be as important to the design of electronic documents as consideration of the physical text structure that gives rise to the cognitive form. The major issue in this context therefore is the extent to which structure is perceived in certain texts and how designers must accommodate this in their products.

6.3 Background to the experiments

As mentioned previously, the author was a participant in a team designing a hypertext database of journal articles. It was obvious from the literature, from theoretical considerations and from the evidence of the BLEND and ADONIS systems that straight reproduction of the paper format was unlikely to prove acceptable in usability terms, even if more advanced manipulation facilities had emerged since either of those systems were specified.[4]

The question of how such texts might be structured to maximise usability was of direct concern to this design project. Furthermore, the journal usage study cited earlier had shown that typical usage of this text type involved jumps and non-serial access routines for which a model of an article's typical structure was a useful guide. Thus there were two apparently conflicting requirements: support non-serial access while retaining a supposedly useful linear structure.

The investigations reported here marked an attempt to clarify these issues in the context of the database under design. The major issue was to identify the relevance of the article superstructure to the database in question though a secondary one was to examine the relevance of this concept to the more general work of electronic text design. Both of these studies are the sole work of the present author.

---

[4] Though both systems are relatively recent (particularly ADONIS which has yet to go on full release) the time between specification and delivery of a system can be considerable. The original BLEND specifications were probably drawn up in or around 1980-82; The ADONIS workstation referred to in this thesis was formally specified in 1986 although discussions about its exact form and content commenced in 1980. The current database referred to here was specified in 1989 (though planning started in 1988). Such differences may appear short but are in fact a long time in information technology terms where the state of the art changes rapidly.

6.4 Overview of the Experiments

The specific aims of this experimental work were to identify the extent to which readers possessed an accurate superstructural representation or model of a typical academic article and to examine how it might be affected by screen presentation.

6.5 Experiment 1

6.5.1 Rationale

If readers possess a model of the how typical articles are structured then they should be able to use this to form whole articles out of isolated chunks of text. They might still be able to perform this without such a mental representation if headings and other cues in the text such as referential continuity are present. The present study examined this suggestion by presenting subjects with cut-up articles and requiring them to piece the articles together. To limit the influence of referential continuity cues, every second paragraph was removed and subjects performed this task on texts with and without the presence of headings.

6.5.2 Subjects

Twelve subjects participated in this experiment (six male, six female). Ages ranged from 21 to 35 (mean=29) years. All were professional researchers experienced in the use of academic articles.

6.5.3 Texts

Two articles were selected from one journal in a field of relevance to the researchers. The articles were matched approximately for size and number of paragraphs, presence of figures and tables, and conformation to the single experiment report style. Though roughly in the area of interest to the researchers they were also selected so as to be unlikely to have been read by these researchers. This was subsequently confirmed during the trial.

The rules for removing paragraphs were not formal. Where possible, every second paragraph was removed but if this left only very large or very small paragraphs (i.e. greater than 20 lines or less than 5 lines respectively) some adjustments were made and experimenter discretion was employed to retain comparability between texts. Every second table and figure was removed from each text. Selected paragraphs, headings (without numbers), tables and figures were pasted to pieces of card to aid physical manipulation.

6.5.4 Design

A repeated measures design was employed such that each subject assembled both texts, one with headings, one without. Order of texts and presence/absence of headings per text were counter-balanced to avoid any systematic order effects.

6.5.5 Procedure

Subjects were run individually in an experimental room at HUSAT. The experimenter explained the task and answered any questions from the subjects. They were told to avoid reading every word in the text if possible and to concentrate on assembling an ordered article as quickly as they could. The text was presented in a jumbled order on the desktop.

After the first text had been assembled to the subject's satisfaction, subjects were asked to move to another desk and write down a brief summary of what they thought the article was about. This enabled the experimenter to score their performance on the first assembly task and prepare the second text. The instructions were then repeated and the subject proceeded to assemble this article. After completion the subject again went to the other desk and wrote a brief summary of their impressions of the article's content. Upon completion a brief discussion of the experiment ensued covering any points the subject wished to raise.

6.6 Results of Experiment 1

6.6.1 Accuracy scores

In the first instance data were scored by noting the relative position of each text chunk in a subject's assembled text and comparing it with its correct position. This gave a measure of the absolute accuracy of assembly. Not surprisingly, no subject manifested a high degree of absolute accuracy, mean rate was 16.7% i.e. approximately five correct placements per 30 paragraph task. A repeated measures t-test revealed no significant effect for headings ($t=0.31$, $df=11$, $p>0.7$).

Despite the low levels of absolute accuracy it was clear that subjects were imposing a structure on the article of the form Introduction/Method/ Results/Discussion (hereafter referred to as the IMRD format). Indeed all subjects assembled the article around this format. Analysing their assemblies in these terms (i.e., awarding a point for each paragraph correctly placed within general category) it was clear that much higher general accuracy

levels were present (mean accuracy rate=82.58%). Table 6.1 presents the individual error scores for this broader analysis.

These data indicate that subjects can predict location of isolated paragraphs of text in their correct general sections with high levels of accuracy. Once more, the effect of headings was assessed using a related samples t-test and this revealed no significant difference (t=1.6, df=11, p >0.1).

Subject	Headings	No Headings
1	14	6
2	1	6
3	0	10
4	5	7
5	2	3
6	2	10
7	6	2
8	0	3
9	3	8
10	5	8
11	2	7
12	4	2
mean	3.67	6
sd	3.8	2.89

Table 6.1 Error scores per subject in broader classification

6.6.2 Speed

All times to completion were recorded in seconds. They are presented in table 6.2. Times in the no headings condition were slightly faster than those in the headings present condition. Both distributions were slightly negatively skewed (no headings= -0.22; headings=-0.97) indicating that the majority of extreme scores are below the mean in both conditions (i.e. some individuals were much faster than the majority). A related samples t-test between conditions showed an almost significant difference at the 5 per cent level (t=2.07, df=11, p<.065). However such a difference was expected as the no headings

conditions always involved slightly fewer pieces of text than the headings present condition.

Subject	Headings	No Headings
1	904	922
2	416	446
3	893	711
4	507	594
5	711	731
6	891	892
7	909	699
8	629	479
9	841	631
10	898	892
11	902	721
12	900	892
Mean	783	718
sd	175	162

Table 6.2 Time taken per subject by condition

6.6.3 Error types

Apart from the general accuracy levels observed, it was interesting to note the type of mistakes made by subjects in this task. Three basic errors can be identified in the present data:

(i) Secondary heading placement
(ii) Figure and table placement
(iii) Introduction/Discussion distinction

The most obvious problem occurred with the secondary headings. Primary headings (Introduction etc.) were easily placed but these are relatively standard, secondary headings tend to be unique to the article, reflecting the author's views of a section's contents. For example, a heading such as "The Effect of Display Size" might fit logically into the results section when read in context but taken as an isolated piece of text could as easily be a heading in the Introduction or Discussion sections of an academic article.

Figures and tables posed problems in terms of absolute accuracy too, although subjects usually placed these in the correct section. This is not too difficult to explain, their

occurrence in articles of this form is rare outside of the results section. Non-graph/numerical types might pose more of a problem but even they are unlikely to occur in Introduction/Discussion sections.

A common error was the confusion of Introduction and Discussion paragraphs. All subjects made this mistake at least once. In terms of the type of text usually found in these sections this is understandable. Both contain general text with references to other related work, a form atypical of other sections. Thus while it is easy to identify isolated paragraphs as belonging to these sections, it is less easy to distinguish correctly between them.

6.6.4 Awareness of text's contents

All subjects were required to describe briefly the contents of the text they had just assembled. Of the 12 subjects, 10 remarked that they had little memory of the text and had not read it for comprehension. As a result they claimed not to be able to write very much. While it is interesting that they could assemble the text without reading it for comprehension purposes, all subjects were capable of providing a rough sketch of the article. Typically they accurately reported the subject matter, that it was an experimental paper, the design or analysis, and its broad aims. In some cases parts of the results or their implications were grasped.

There were inaccuracies however and most of the written reports were in the form of keywords or short phrases suggesting little attempt to grasp the development of the argument within the text. This supports the claims of subjects not to have read for comprehension.

6.6.5 Conclusions from Experiment 1

It is clear from these findings that readers of academic articles possess some form of mental representation for the text's typical structure that allows them to predict accurately a paragraph's location. In the present case this representation seems to be of the form IMRD and a quickly read paragraph can be placed in this framework with approximately 80% accuracy. Problems occur with secondary headings, absolute placement of items within the framework and distinguishing between introduction and discussion text.

6.7 Experiment 2

6.7.1 Rationale

It is clear from experiment 1 that readers do possess a model or mental representation of a text's structure independent of its semantic content. However, much work in this area of electronic text has shown that when text is presented on screen many of the findings from the paper domain cease to hold. The present study therefore set out to examine the ability of readers to predict location by applying the superstructural representation of an article to information on screen.

6.7.2 Subjects

Eight subjects (4 male/4 female) participated in this study. Ages ranged from 21 to 41 (mean=32) years. As before, all were experienced users of academic articles. Three of the subjects had participated in the previous study but given the seven week break between the studies and the use of different texts and experimental procedure this was not seen as a source of contamination. All were habitual users of Apple Macintosh computers.

6.7.3 Texts

As before two similar articles conforming to the criteria described above were selected from a relevant journal. This time only text was selected (i.e., no figures, tables or headings were used), five paragraphs from each major section resulting in 20 paragraphs per text. These were presented in a randomised order which was consistent between media.

6.7.4 Design

A repeated measures design was employed with order of presentation (paper and screen) and text counterbalanced to avoid any systematic ordering effects.

6.7.5 Procedure

Subjects were run in an experimental room at HUSAT. The experimenter explained that they had to read two series of 20 paragraphs and identify the probable location of each in terms of the major sections Introduction/ Method/Results/ Discussion. To do this they marked I, M, R,or D on an answering sheet provided. They were told to perform this task as fast as they could.

In the screen condition paragraphs were presented mid-screen as black text on a totally white background using HyperCard on an Apple Macintosh Plus. The only other information present was the number of the paragraph (1 to 20) in the top right corner and a "button" in the lower centre of screen facilitating movement to the next card. In the paper condition paragraphs were presented on 20 sheets of paper printed from this HyperCard stack, of similar size to the screen and stapled together in the top left corner. They contained identical information except for the "button".

Subjects were allowed to familiarise themselves with the task and the software (usage of which only required them to press the mouse button) using example texts and the experiment commenced when they expressed confidence with both. A rest period of approximately two minutes occurred between the two trials.

6.8 Results of Experiment 2

6.8.1 Speed

Time taken to complete each trial was recorded in seconds and these are shown in table 6.3. As this demonstrates, mean performance time with paper was faster than with screen presented text. A related samples t-test indicated that this difference was significant at the 2 per cent level ($t=3.16$, $df=7$, $p<.02$).

Subject	Paper	Screen
1	119	210
2	121	193
3	325	388
4	365	370
5	240	301
6	255	330
7	130	160
8	315	285
mean	234	280
sd	99	84

Table 6.3. Time to complete tasks per condition

6.8.2 Accuracy

The number of errors made during each trial by each subject is shown in table 6.4. This demonstrates that the mean number of errors per subject is similar for each presentation medium although there is greater but non-significant variance among the scores in the screen condition ($F=0.11$, $df=15$, $p>.7$). Interestingly, six of the eight subjects performed better or as well with the electronic text. A related samples t-test showed no significant difference however ($t=0.32$, $df=7$, $p>.7$).

Subject	Paper	Screen
1	4	2
2	4	0
3	5	1
4	4	9
5	4	4
6	3	7
7	4	3
8	3	2
mean	3.88	3.50
sd	0.64	3.07

Table 6.4. Number of errors made by subjects per condition

Overall accuracy levels are similar to experiment 1, 81.55% for combined conditions, 80.6% for paper alone, 82.5% for screen alone, confirming the earlier finding that the ability to predict location on the basis of limited information is highly developed for experienced readers of this text type.

6.8.3 Error types

The absence of headings and figures/tables in the present study made quantifiable analysis of the error types easier and qualitative analysis less informative. Twelve possible errors could be made (No. of categories x No. of incorrect categories per item). In total, 59 errors were made. These are summarised in table 6.5.

Error Type [Item] to [Incorrect place]	Frequency	%
Introduction to Method	1	1.69
Introduction to Results	1	1.69
Introduction to Discussion	16	27.12
Method to Introduction	3	5.08
Method to Results	5	8.48
Method to Discussion	0	0
Results to Introduction	0	0
Results to Method	5	8.48
Results to Discussion	12	20.34
Discussion to Introduction	7	11.87
Discussion to Method	3	5.08
Discussion to Results	6	10.17

Table 6.5. Error type and frequency expressed as a % of total errors

As before the greatest difficulty subjects had was distinguishing between the Introduction and Discussion sections, these accounting for almost 40% of errors. Inability to distinguish between the Results and Discussion sections accounted for 30% of errors while the Method and Results distinction proved the stumbling block in 17% of cases.

6.8.4 Conclusions from Experiment 2

Readers' models of this text's structure allows them to predict accurately the general location of paragraphs even when presented on screen. Though significantly faster with paper there were no differences between the media in terms of accuracy. As before, greatest difficulties occurred in distinguishing between Introduction and Discussion sections.

6.9 General Discussion

It is clear from these findings that readers who are experienced in the use of this text type possess a superstructure or model of it which enables them to predict with high levels of accuracy where information is located. In the case of the text type analysed here, the academic experimental article, this superstructure is of the form: Introduction, Method, Results, Discussion and readers can place paragraphs correctly within this framework with approximately 80% accuracy under time pressure.

The existence of this superstructure probably results from the relatively standard form of such articles. There are few published accounts of experimental work in this (and other)

disciplines that do not conform to this type. Obviously, frequent readers of this text type would acquire an awareness of such a form over time.

However, it is also worth noting that the classic IMRD structure acts as a framework for or model of the scientific process itself. Research usually takes the form of examining the current literature to formulate a hypothesis for investigation, designing an experimental procedure to test this hypothesis, gathering and analysing data, and finally examining the results in the light of other work. Each of these activities has its parallel in the resulting description i.e., the experimental report. Generations of undergraduates are taught this model of investigation and reportage (even if it is, as Medawar (1964) stated, more a reflection of what scientists would like readers to think they have done rather than what they actually did!) so it is not surprising to find superstructures for this emerging. In a very real sense therefore, text structures can reflect conventions and standards of behaviour and cognition as argued by van Dijk (1980) and van Dijk and Kintsch (1983).

It must be recognised however that an alternative interpretation cannot be entirely ruled out. It is possible that the control on referential continuity was not strong enough and that the successful piecing together of articles in experiment 1 may have been helped by cues that could not be completely removed by using only every second paragraph. It is the author's contention that such factors did not play a significant role in subjects' performances, not least because of the subjects' reports that they did not read with the intention of seeking reference cues between paragraphs and the fact that when reading isolated paragraphs one at a time in a randomised order in experiment 2 (thus guarding against any reasonable use of such cues), general accuracy levels were equally high. However, further credence in the theory of supertructures would be gained if a control group of non-experienced article readers was examined and found to manifest significantly less accurate scores. Such a study would also overcome any possible demand effect that may have been present as a result of any subject's possible knowledge of van Dijk and Kintsch's work. This is a potential weakness of the present design that should be addressed in any future work.

Furthermore, the number of subjects used in these studies was relatively small. It is possible that some of the non-significant differences, particularly those concerning the effect of headings on ability to piece together an article (experiment 1) and the possible existence of a speed/accuracy trade-off (experiment 2) may hide real issues that would have been uncovered had more subjects been used. The reason for not using greater sample sizes in the present work stems from the demands on the design team for a prototype system and the subsequent pressing need for quick answers to questions on document structure. In an ideal world such constraints would not operate and large sample sizes could always be employed. While this is

an area for improvement in subsequent research, the basic findings of the present studies do appear robust.

Regardless of any hypothetical cognitive representations underlying text usage, what is interesting from a human factors perspective is the high degree of accuracy shown by all subjects in these experiments. From a rapid scan of the available text they can deduce the most likely location of that part in the whole and by extension, what is likely to precede, accompany and follow it. The results of experiment 2 clearly demonstrate that this representation holds also for screen presented text.

If designers are to consider seriously alternative structures for electronic or hypertext versions then they would need to overcome this acquired processing tendency of experienced readers. This is an all too unlikely occurrence given the embedded nature of this representational structure in the minds of readers, the teaching of the scientific process and the communication format of scientists.

Thus a hypertext journal article would need to retain the broad structure of the paper versions if it is to be immediately usable. The superstructure or model should be used to enhance the reader's ability to navigate, reportedly the major problem for many hypertext users (see e.g., Edwards and Hardman, 1989). For example, keeping the major headings and their standard order as the "backbone" of the text would facilitate rapid exploration of required sections and narrow the search space for information location. Combined with the rapid access facilities of hypertext, such a format could result in the development of an electronic text that would ideally suit several of the reading tasks common to this text type identified in the journal usage study. This course of action was decided upon by the design team of the hypertext journal database discussed earlier.

One other interesting aspect of these studies that warrants discussion is the significant speed deficit of 17% for screen-presented text found in experiment 2. There are several possible sources of this. Firstly, when using the HyperCard stack, subjects manipulated the text by positioning the mouse on the "Next" button and pressing once. It is possible that between manipulating text in this manner and writing the answer by hand on paper more demands were being placed on the subject than when using paper alone. Given the usual habit of subjects to control a mouse with their preferred (i.e., writing) hand this might have slowed them down. Against this however it must be remembered that subjects needed only to position the mouse on the first card, after which each button depression left the cursor positioned correctly on the next card. Unfortunately, this aspect of performance was not monitored during the experiment.

A second explanation is the image quality hypothesis discussed in chapter two. According to current research, reading from screens is approximately 20-30% slower than reading from paper due to the poorer image quality of screens (Gould *et al.*, 1987a). The only screens that seem capable of matching the image quality of paper are very high resolution with black text on white backgrounds using anti-aliased characters (see e.g. Gould *et al.*, 1987b). In the present situation, the screen was a standard one, the text, though black on white, was not presented in an ideal screen font such as Geneva but in New York (10 point), which is a screen optimised version of a paper optimised font, Times which closely matched the paper font used (the reason for this was to try and retain as much comparability between the media as possible). It is possible therefore that image quality was responsible for the speed deficit which in this case showed screens to be almost 20% slower. Against this argument however, it must be stated that the amount of text being read was very limited, and image quality effects should have been very subtle. It is probable that both explanations, reflecting two levels of the classification in chapter two (manipulation facilities and visual ergonomics) are contributory factors to the observed significant difference.

It is perhaps to be expected that the format of IMRD is very familiar to readers of academic articles. However, the main point of these studies was not to confirm this fact but to examine the extent to which the perception of structure influenced readers organisation of the text. The ease and speed with which these subjects arranged the material or predicted its location suggests that for this text type at least it is a very potent aid to organisation. Other text types are likely to have less clear superstructures and in these cases, alternative structures for hypertext versions should be investigated. What seems likely though is that readers do acquire some knowledge of structure for all texts, and that it increases with experience in using that text type. In use, it is likely to combine with spatial memory for layout (Rothkopf, 1971) to form a mental map of the text being read, facilitating searching and browsing of the material. Such issues must be addressed by the designer of any text presentation system if usability is to be ensured.

6.10 Conclusions and the way forward

The issue of text structure is relevant to electronic text design and needs to be considered in its development. In the case of academic journals experienced readers utilise a superstructural representation based on the Introduction, Method, Results and Discussion of a paper and can use this to predict the likely location of items within a document. For the

hypertext journal database being developed at HUSAT these results acted as direct input to the design, suggesting how the individual articles should be structured to aid readers.

In terms of the descriptive framework that is at the heart of this thesis it would seem that the concept of structure is important. Though the present studies concentrated on one specific text type there is no reason to doubt that the notion of superstructural representation extends to other documents (although this is an obvious area for further work).

In conjunction with the earlier work on journal usage a significant amount of relevant data had been now fed back to the design team. This enabled the initial specification to be elaborated and the development of a prototype to begin. This would therefore seem a suitable point to take stock of the type of information provided and outline the framework of description that emerges. This is done in the next chapter.

CHAPTER 7

A FRAMEWORK FOR THE DESIGN OF ELECTRONIC TEXTS[1]

7.1 Introduction

The output from the studies reported in the previous chapters has provided useful information to the team designing a hypertext database. However, these studies were carried out in the first instance to provide input to a series of specific design problems. Though these have been successful, the issue of generalisation is paramount. In other words, in what sense has this work taken us closer to the stated goal of a description of the reading process suitable for the analysis of general electronic text design? It is this question that is addressed in the present chapter.

7.2 Relevance of the knowledge gained to designers of an electronic text

For the design team building a hypertext database of journals at HUSAT, the information from the studies reported in the previous chapters was fed back in the form of findings and resultant discussions of implications. For example, the typical behaviour of journal users was presented as a written document to which the design team responded and considered the implications for the database under development. The reading style data were seen as new and extremely relevant evidence. The latter aspect in particular was important. There is little or no data in the published literature which offers comparable information on the reading strategies employed by academic article readers. The resultant discussions tended to highlight other issues that were important and worthy of investigation at that point in the design cycle thereby providing the stimuli for further work (details of this design process and the resultant database can be found in McKnight *et al.*, 1990b).

Although most of the data presented so far concentrate on journal articles, there seems little reason to doubt that the approach and findings are generalisable to other texts. The repertory grid study showed how any text can be described by the same criteria, the usage studies demonstrated that two very distinct texts and their usage are amenable to analysis in these terms, and the concept of structure is almost certainly not unique to any one or few text types. That the design team found these studies more pertinent to their needs than any of the existing literature however is a more important issue at this

---

[1] The concepts and ideas presented in this chapter and the resulting framework are the sole work of the present author.

time as it highlights the form that a suitable framework could take and the level of description required by designers. To understand how such knowledge might be useful however it is important to view human factors knowledge in the broader context of the design process. In the following sections a brief summary of the application of such knowledge in contemporary systems design is presented before the framework encapsulating the output from the earlier studies is presented.

7.3 Theory versus Empiricism: the role of human factors in the design process

Ergonomics is often criticised for being piecemeal rather than coherent, evaluative rather than predictive and addressing specific issues in a way that leaves little scope for generalisation of findings (Chapanis, 1988). This is reflected in the ergonomics practices carried out in systems design where human factors are often considered at a stage too late to effect better designs i.e., the human factors specialist is seen as having a primary role in testing instantiated designs rather than influencing the initial specifications. Yet, given the much sought after opportunity to become involved in the early specification phase, ergonomists suffer from a lack of the conceptual tools and techniques necessary to overcome resistance from engineers seeking better inputs than vague or inflexible guidelines.

The standard philosophy underlying much human factors work is that of iterative user-centred design (see e.g., Norman and Draper, 1986) involving the development of prototypes and their subsequent evaluation, leading to further prototyping and so forth. While such an approach, properly executed, makes the development of usable technology more likely, it is a non-optimum process which can prove extremely expensive in terms of time and resources. Few design companies therefore are willing to invest the necessary effort to iterate through several cycles (Hannigan and Herring, 1987). This had led to the attempted reduction in the number of iterations needed and a move to bring human factors inputs into the design process earlier (Bertaggia *et al.*, 1987, Catterall *et al.*, 1989). Current emphasis is on rapid prototyping facilities which allow designers to mock-up disposable simulations quickly and cheaply. These can help but even then, the quality of the original prototype is dictated by the accuracy of the designer's conceptualisation of the intended users. This is an area that requires important human factors work.

According to Card *et al.* (1983), an applied science of the user that is theory based rather than empirical, using a common framework to provide insight and integration, is

what is needed to aid designers. By this they mean that analytical techniques that do not require any empirical input could be used at the earliest stages of the process. As noted in chapter three, they propose a constrained version of cognitive psychology, with its emphasis on the information processing aspects of humans as the most suitable vehicle for this science and argue that if it is to have an impact, such a science must be based on task analysis, calculation and approximation which would lead to quantitative performance models of users. The role of such a framework would be to encapsulate some relevant knowledge of the user (often termed a user model, user typology or user stereotype) and/or the task (similarly termed task model etc. by some writers) that could provide guidance to the designer specifying the system. According to Card *et al.* the true role of an applied psychology is to provide such performance models in quantitative form for designers.

In reality, the apparent extremes of frequent empirical iterations and formal theory-led designs are merely opposite ends of a continuum; ends in which few human factors practitioners (or designers) permanently reside. More common is a mixed approach linking empiricism to theory and vice-versa, with a bias towards empiricism due to the perceived lack of relevant (i.e., applicable) theoretical models. The mixed approach is probably inevitable as both extremes are impossible to implement absolutely. All observation is theory impregnated according to most contemporary philosophers of science (see e.g., Chalmers, 1976). Thus any system is going to be coloured by assumptions about the user, however implicit, when it is being developed prior to testing.[2] The empirical route to design can therefore in no way claim to reject totally theoretical perspectives of the user in favour of experimental facts. However, complete theories of human performance in HCI (or anywhere else for that matter) are non-existent and any design based on theoretical models alone must be evaluated by empirical means to ascertain its true level of usability.

Therefore, a practical goal for frameworks and models in HCI is to guide the derivation of suitable initial designs which, by virtue of their accuracy, reduce the number of iterations required before an ultimately acceptable design is achieved. Evaluations would subsequently act as confirmation or rejection of the design (or parts thereof) and if the latter, lead to refinement of both the resulting system and the theoretical framework underpinning it. The value of frameworks or models therefore lies both in their ability to reduce iterations and to be modified, if necessary, in the light of data from users and subsequently applied to other designs.

---

[2] In philosophical terms one could argue that even the design of the experiment to test a system reflects underlying belief systems and is therefore theory impregnated.

7.4 Frameworks and Models: a clarification of terminology

The terms framework and model tend to be used interchangeably in the literature, with model being the dominant descriptive term for such theoretical views of users. However, for purposes of clarity, the present work will draw a distinction between them in the manner described previously by Whitefield (1989). He describes a framework as a generic representation of the important aspects of the user and a model as a specific representation of those aspects in relation to a task. In this sense a framework provides the perspective of the user (i.e., reader) for all instances of interest while the model is derived according to the interaction of particular task demands and user. In these terms, the GOMS approach of Card *et al.* for example, may be described as a *framework* from which specific performance *models* are derived. Whitefield (1989) uses the term *framework* to apply to the "blackboard" architecture of AI theorists (e.g. Hayes-Roth 1983) which he borrows to *model* specific instances of problem-solving behaviour in engineering design.

In the present case a framework is proposed of the generic aspects of the reader which it is hoped will support the derivation of more specific models of reader-text interaction for particular tasks. Several important criteria impinge on any proposed framework beyond the obvious one of utility. First, it must be accurate. This is not to say that it must offer a precise picture of the user and text interaction being supported but what it offers should be correct in the sense that it describes real factors or aspects that influence the reading process. Second, it must be non-complex. Invoking psychological descriptors or cognitive structures in a form suitable for non-specialists to use and apply is a difficult but necessary part of a good framework. Third, it must be suitably generic to be of relevance to more than one application. Just as the the reading process covers a myriad of texts and tasks, designers should be able to utilise the framework describing this process for guidance on the design of more than one text system. Finally it should be modifiable. This does not mean that it must be altered every time it is used but that it should be capable of being adjusted in the light of feedback. The following descriptive framework is an attempt to satisfy all four criteria. The next section outlines the framework in detail.

7.5. The proposed framework

The framework is intended to be an approximate representation of the human cognitions and behaviours central to the reading process that are employed in the interaction between reader and document. It consists of four interactive elements that reflect the

primary components of the reading situation at different phases. These elements represent the major factors deemed to be important to reading on the basis of the work in the preceding sections. They are:

1. A Task Processor (TP) that deals with the reader's needs and uses for the material;
2. An Information Model (IM) that provides a model of the information space;
3. A set of manipulation skills and facilities (MSF) that support physical use of the material;
4. A Serial Reading Processor (SRP) that represents the cognitive and perceptual processing involved in reading words and sentences.

These are not isolated variables but interrelated components reflecting the cognitive, motor and perceptual aspects of reading. In other words, according to this framework, reading is not a matter of merely scanning words on a page or acquiring and/or applying a representational model of the text's structure but a product of both these activities in conjunction with manipulating the document or information space and defining and achieving goals. So for example, a reader recognises an information need, formulates a method of resolving this need, samples the document or information space appropriately applying his model of its structure, manipulates it physically as required and then literally perceives (in the narrow psychological sense) words on text until the necessary information is obtained. Obviously this is a very simple picture of the reading process, other more complex scenarios are possible such as the revision of one's reading goal in the light of new information or modifying one's initial information models to take account of new details and so forth. The point here is that regardless of the precise scenario, the elements described here should cover all important aspects from the point of view of the text designer. They are the building blocks of the activity described as reading which can be combined in numerous permutations. Each of these elements and their various interactions are described in more detail the following sections.

7.5.1 The Task Processor (TP)

The notion of the reading task as the crucial factor in understanding text use provides a sound basis from which electronic text design can be investigated. Readers interact with texts purposively, to obtain information, to understand, to learn etc. To do this they must allocate cognitive resources to some form of task processor that decides what it is

they want to get out of the text and also reviews their progress and, if necessary, revises the task.

This notion of intentionality in reading gives rise to the idea of planning in the reader's mind. The extent to which such plans exist is theoretically debatable but it seems reasonable to infer that the extent to which reading is a goal-driven behaviour some level of planning how to interact with the information source must occur.

From the task analyses and repertory grid studies carried out earlier it seems that such planning is relatively gross, taking the form of such intentions as "go to the index, look for a relevant item and enter the text to locate the answer to my query" or "to find out what statistical tests were used go to the results section and look for a specific description". However, they can be much vaguer than these two examples which probably represent highly specified plans of interaction with the text. Reading an academic article to comprehend the full contents seems to be much less specifiable, the reader is more likely to formulate a plan such as "read it from the start to the finish, skip any irrelevant or trivial bits, and if it gets too difficult jump on or leave it". Such a plan may be modified as the reading task develops e.g., the reader may decide that he needs to re-read a section several times, or may decide that he can comprehend it only by not reading it all. In this sense planning becomes more like a situated action (Suchman, 1988) where the reader's plans are shaped by the context of the on-going action and are not fully specifiable in advance.

Whatever the precise nature of the plan it seems appropriate to posit a task processor that generates some method for dealing with the document or documents under consideration. Thus the framework must accommodate such activity in order to focus designers' attention on such reader-text interaction charactersitics.

7.5.2 The Information Model (IM)

Readers possess (from experience), acquire (while using) and utilise a representation of the document's structure that may be termed a mental model of the text or information space. Such models allow readers to identify likely locations for information within the document, to predict the typical contents of a document, to know the level of detail likely to be found and to appreciate the similarities between documents etc. The journal and manual usage studies as well as the experiments in the previous chapter highlighted the existence of such models and it is postulated here that such representations exist for all commonly used text types.

However, it is worth making a distinction here between what Brewer (1987) terms "global" and "instantiated" schemata with regards such mental models. In the present context a global schema consists of a representation of how a typical text type is organised e.g., an experimental article is typically made up of Introduction, Method, Results and Discussion sections, or a newspaper is made up of a series of articles covering a range of topics grouped into sections on politics, sport, finance etc. These are the type of structural representations that are general and exist independently of any specific document (though of course they only emerge over time after frequent interactions with many documents).

An instantiated schema consists of an embodiment of the generic model based on exposure to a specific text, e.g., noting that the particular article one is reading has a very short introduction or there is a reference to "Bloggs (1982)" on the top of a right hand page containing a figure. In other words, when a reader interacts with a text, the original structural model of the text type becomes fleshed out with specific details of the particular text being read. The terms "global" and "instantiated schema" are overly technical however therefore this distinction will be referred to here more simply as the difference between a model (which is generic) and a map (which is specific). In these terms readers can be said to form mental maps of particular texts as they use them, models help them in this but are not themselves essential for map formation (i.e., it is assumed that a reader can form a detailed map of a document without having been exposed to similar types of text before). In this way, frequent map formation with a document type can be seen as supporting model formation of that document type's generic structure.

In use, the information model helps the reader to organise the text's contents by fitting it into a meaningful structure and thus guards against navigational difficulties by providing context i.e., it supports the formation of a mental map of the information space. Thus what is initially a model becomes, with use, a map of a specific text. Where no model exists in advance, a map can be formed directly. The point at which a model becomes a map is difficult to quantify and probably not pertinent to present needs. After all, knowing one structural detail about a specific text hardly conveys the idea of a map as the term is typically understood. Therefore the term information model is retained for general use in the framework, map will only be employed when discussing readers detailed knowledge of a specific text's structure.

7.5.3 Manipulation Skills and Facilities (MSF)

Readers must be able to manipulate text. This simple statement hides innumerable complex issues in the design of electronic text. Except for very short passages, documents cover more than one page or screen and the reader needs to be able to physically alter his view of the material. Even in such situations he will still need to locate and "open" the text for reading, actions which clearly involve manipulation. With paper, such skills are acquired early by readers and are largely transferable from one text form to another. If you can manipulate a paperback novel you will have few difficulties with a textbook and so forth, although there are obvious exceptions in the paper domain and the ability to easily manipulate broadsheet newspapers in confined spaces is a specific skill that is relatively unique to that text form.[3]

However, such paper-based skills are potentially limited in terms of what you can do with the text. Most readers are skilled in using their fingers to keep pages of interest available while searching elsewhere in the document or flicking through pages of text at just the right speed to scan for a particular section, but beyond these actions, manipulation of documents becomes difficult. When one then considers manipulation of multiple documents these limitations are exacerbated.

Electronic text is awkward to manipulate by means of scrolling or paging alone but the advent of hypertext with its associated "point and click" facilities has eased this somewhat. However, the immediacy of interaction with electronic text is less than it is with paper by virtue of the microprocessor interface between reader and information on screens. Furthermore, the lack of standards in current electronic information systems means that acquiring the skills to manipulate documents on one system will not necessarily be of any use for manipulating texts on another. Obviously electronic text systems afford more sophisticated manipulations such as searching which can prove particularly useful for certain tasks and render otherwise daunting tasks (such as locating every reference to a certain topic in the complete works of Shakespeare) now manageable in minutes rather than days. Yet it has been shown that such facilities are not always a guarantee of accurate performance.

The various advantages and disadvantages of manipulation facilities on screens have been presented in detail in chapter two. Ultimately, the goal is to design transparent

---

[3] Anybody who has witnessed the elegant manipulations of individuals reading broadsheet newspapers such as The Times or The Guardian while travelling on the London Underground will appreciate the skill factor involved.

manipulation facilities that free the reader's processing capacity for task completion. Slow or awkward manipulations are certain to prove disruptive to the reading process. The framework raises these issues as essential parts of the reading process and therefore important ones for designers to consider in the development of electronic text.

7.5.4 Serial Reading Processor (SRP)

The final element of the framework is the serial reading processor. It is proposed that this is the module that actually processes the images from the document and carries out the activities most typically described as "reading" in the psychological literature (e.g., Just and Carpenter 1980). Thus eye movements, fixations, letter/word recognition and other perceptual, linguistic and (low-level) cognitive functions involved in extracting meaning from the textual image are properly located at this level.

The prefix "serial" on the reading processor emphasises the perspective that at this level of cognition reading generally occurs in a serial fashion. That is not to say that texts are accessed and used serially (which they are obviously not according to the evidence presented in this thesis) but that at the level of engagement between the eyes and the document reading is for the most part serial. Obviously regressions occur and people jump about from one part of a page to another without physically manipulating the document but even then, this requires readers' attention to focus briefly on their model of the text and so changes the cognitive processes that are being dealt with. At the level detailed here, information extraction from a document relies on the reader serially processing letters, words and sentences.

The question of how serial reading is will not be discussed further as it is not pertinent to the present thesis. Decades of psychological investigation have been spent looking at the question of how humans read and some of the conclusions drawn from this work have been discussed in chapters two and three. Present emphasis dictates that the findings on eye movements, reading speeds, letter and word recognition etc. are considered sound but are of relevance here only to the extent that reading electronic text is influenced by or alters these aspects of the process. An obvious example of how issues at this level affect electronic text design is to be found in the image quality work of Gould *et al.* (1987b).

7.6 Interactions between the elements

So far, the basic components of the framework have been described. These reflect the

human aspects of performance during the reading process and are therefore the elements that seem pertinent to electronic design. A schematic representation of the framework is presented in figure 7.1.

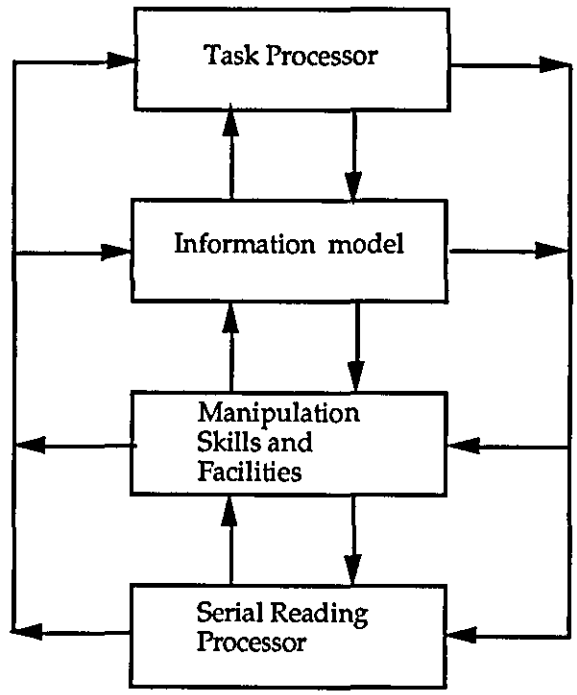


Fig. 7.1 The framework for describing reading

According to the framework there are 12 possible interactions between these elements that can occur. These will be described individually.

• Task Processor to Information Model

When a task is formulated the reader usually interprets or mediates its formulation and expectation of its outcome in terms of his model of the information space. For example, if the task is "Find the reference in the text to Bloggs", the reader applies the model to narrow the search space and produce an inference such as "It is more likely to be in the section labelled 'Related Work' than in 'Results'." This is a natural and rapid occurrence.

• Task Processor to Manipulation Skills and Facilities

Where an information model does not exist (when this is a reader's first exposure to a

text type for example) the reader, upon formulating a task may proceed to manipulate the text without any knowledge of its layout or contents. This may be as simple as opening a book (or electronic file) with no better intention than reading it from start to finish until the target information is located. In terms of the framework this is a case of direct interaction between the TP and the MSF elements. It implies that absence of a model at the outset does not prevent text usage.

• Task Processor to Serial Reading Processor

In cases where the text is short and available e.g. a single page memo on one's desktop, even manipulation facilities may be unnecessary. Similarly during a particular sub-task of a larger one, a reader may engage only TP and SRP to perform that sub-task e.g., locate a word in a paragraph currently on screen or on the open page. This example highlights the fact that it is possible to read without engaging either the IM or MLS elements.

• Information Model to Task Processor

A model of the information one is dealing with can influence the type of task one tries to perform with it and aid accurate specification of that task. For example, if a reader's goal is to find out about particular theories of child development, a model of the text (e.g., an introductory psychology textbook) could suggest that the book was inappropriate but it might offer suggestions for further reading. On the other hand, such an interaction between IM and TP elements could occur where, after reading the text for several minutes the reader's model might indicate that the text is unlikely to contain the form of information required and therefore the task needs to be re-specified.

• Information Model to Manipulation Skills and Facilities

The interaction between these two elements in this direction is likely to be of the form: model directing manipulation e.g., the information being sought is at the end of the text therefore page or scroll to the last chapter. Such rapid interactions should characterise many reading situations.

• Information Model to Serial Reading Processor

Again, this is only likely to occur for particular tasks and very short texts. An example might be identifying the sender of a one page letter. In this case one's model of the

letter form suggests that an address may be provided at the top of the page or a signature will be present at the bottom. Once the letter has been opened and the page unfolded further manipulation activities can be by-passed.

• Manipulation Skills and Facilities to Serial Reading Processor

Once the task and model aspects have been applied and the text is of the type that will require manipulation, an interaction between the MSF and SRP elements occurs. A simple example is the reader turning a page to allow reading to commence.

• Manipulation Skills and Facilities to Information Model

Such an interaction might be expected to result when, faced with an unfamiliar text, the reader manipulates it and induces the formation of a primitive information model (the "flick through to see what's in it" approach).

• Manipulation Facilities to Task Processor.

In this instance the information flows directly from the element concerned with text manipulation to the task processor. Though presumably rare, an example might be when a reader finds that he can't search for a term and therefore cannot perform the task as originally envisaged.

• Serial Reading Processor to Manipulation Skills and Facilities

Where the reading processor is interrupted by a page break or screen end an interaction with manipulation facilities occurs to facilitate further SRP activity with the text. This is logically distinct from the MSF to SRP activity described earlier which refers to activity occurring prior to any SRP activity.

• Serial Reading Processor to Information Model

The reading processor may interact directly with the information model by providing information to the reader about the contents of a page or the type of material contained in the document at that point thereby supporting the formation of a map, for example noting the occurrence of a particular word or phrase when reading as a potential landmark in the document.

• Serial Reading Processor to Task Processor

The direct interaction between the two extreme elements in the model may occur in this direction when for example, information is read which solves the immediate task or sub-task e.g., if, when searching for a word or phrase in a certain section, the reader perceives it and thereby resolves their immediate task without requiring further manipulation or model activities.

In practice, it is not hypothesised that such neat interactions occur in isolated units. Meaningful engagement with a document is more likely to result in multiple rapid interactions between these various elements. For example, a scenario can be envisaged where, reading an academic article for comprehension, the task processor interacts with the model to identify the best plan for achieving completion. This could involve several TP —> IM and IM—>TP interactions before deciding perhaps to serially read the text from start to finish. If this plan is accepted then manipulation facilities come into play and serial reading commences. The MLS—>SRP interaction and SRP—>MLS interaction may occur iteratively (with occasional SRP—> IM interactions as distinguishing features are noted) until the last page is reached at which point attention passes back ultimately to the TP to consider what to do next.

Also, the speed and the iterative nature of the interaction between these elements is likely to be such that it is difficult to demonstrate empirically the direction of the information flow. In many instances it would be virtually impossible to prove that information went from MSF to IM rather than the other way and so forth. However this does not preclude examination of these elements and their interactions in an attempt to understand better the process of reading from a human factors perspective. The elements reflect the major components of reading that emerged as important from the studies earlier in this thesis and are intended as a broad representation of what occurs during the reading process.

7.7 The framework in comparison to existing theories: qualitative versus quantitative representations.

The framework presented above is a relatively simple representation of those issues found to be of importance to the usability of an electronic text. They are described in this framework qualitatively i.e., their actions are presented in general terms rather than being specified formally. The absence of rules of exclusion/inclusion or numerical values necessitates greater interpretation by an eventual user than a quantitative

framework or model of HCI or reading. This is intentional, a matter of choice (and some necessity given current knowledge) rather than a failure on the author's part to specify further the framework's components. In the present section the case for such a framework is presented by considering it in the light of the usage of typical quantitative models more commonly expounded in this domain.

As outlined earlier in the thesis, traditional psychological models of the reading process are very detailed, postulate the existence of numerous cognitive structures and processes and tend to concentrate on isolated aspects of the reading process such as word recognition, sentence processing or eye movements. It has been argued at length in chapter three that the level of detail provided by such models of human information processing are too low to be applied in HCI and that in the case of reading this severely hampers the development of usable electronic text systems. The form of modelling common to cognitive psychology is mirrored sharply in the attempts of human factors professionals to describe or model human behaviour at the computer interface. As detailed in chapter three, the major research effort has concentrated on developing formal models of a quantitative kind for designers to apply at the specification stage of product design.

Advocates of the quantitative approach cite precision, non-ambiguity of terminology and ability to calculate design trade-offs as major advantages of such models (see e.g., Harrison and Thimbleby, 1990). While this may be true for models such as GOMS or Cognitive Complexity Theory (CCT) (Kieras and Polson, 1985) when used for very specific analyses (and there is little by way of confirmatory evidence of this yet) there are two underlying assumptions in this view which are directly pertinent to the present work.[4] One is that designers find such quantifiable outputs relevant and the other is that the human performance and behaviour one is interested in can be reduced to suitable numeric functions.

From what is known about the way designers work in real-time, theoretical quantitative models seem to have little relevance in their current form (Buckley, 1989, Carroll, 1990). Virtually all successful reports of the application of these models emerge from experimental work in academic rather than industrial environments (see e.g., Polson *et al.*, 1986, Polson *et al.*, 1987). Their proponents might claim that they are useful and reliable but the design community remains unconvinced. This could result from several

[4] There are many criticisms of these formal methods that could be made which are not directly pertinent to this discussion e.g., what of their accuracy? why do independent users of the methods often derive different models of the same problem? or why are they so difficult to use? Critiques of these aspects can be found in Carroll and Campbell (1986),Winograd and Flores (1988) and Sharratt (1987)

reasons not related to the scientific validity of the approach such as the difficulty of applying them (they usually require substantial domain knowledge to use effectively) or their concern with narrow aspects of tasks rather than global user behaviour which renders them more suitable for application after the initial specification rather than before. According to Landauer (1987) such models do not tell a designer how to design a good system in the first place (which is what they want). Instead, they just advance the moment when evaluation can first be carried out to the pre-prototype stage i.e., they are a measurement tool rather than a creative design aid.

The present framework takes such shortcomings as its starting point and is designed to offer a conceptual aid to electronic text design that does not suffer such problems. First, a designer does not require sophisticated knowledge of human cognition or the psychology of reading to comprehend the framework. Obviously detailed psychological work underlies concepts such as information models, task processors etc. but the designer can consider the basic issues without possessing such knowledge. The more knowledge of human cognition that a designer possesses the more critically and usefully he may be able to apply this work but such knowledge is not a prerequisite for use. Second, unlike GOMS or CCT, the present framework does not require the use of a formal language or sequences of rules to support interpretations of likely user behaviour. It is intended only to draw attention to issues such as image quality and information organisation in the first instance so that the designer realises what is important in a design, not to provide a means of calculating design trade-offs in terms of performance times. Third, the framework covers the full range of behaviour described as reading as it impacts on system design, not just a particular subset of it. It is intended to cover reading as it pertains both to proofreading and scanning of lengthy texts for example, or to using textbooks or magazines. Finally, it is suggested that a successful electronic text system is one that addresses all four elements of the framework in its design, therefore a designer can employ it to guide his initial specifications, i.e., it is a design aid more than a measurement tool.

Such explicit qualitative models of human performance are not commonplace in HCI (though implicit ones abound) but the framework is not without precedent. In psychological terms for example, guidelines as simple as "the human is an information processor with five sense channels" may be considered a highly simplified qualitative model. This is a particularly broad representation and as such, is of very little use to designers. Norman (1986) provides a more specific qualitative model of human interaction with systems which consists of seven stages ranging from forming an intention to act to evaluating an outcome in terms of goals sought. Dillon (1987)

presents a qualitative model of user knowledge development in terms of three stages: confusion, rationality and knowledge. Neither of these offer quantitative power to the designer but they do elaborate or make explicit certain "myths" or standard beliefs about users e.g., that they are goal seeking (Norman's model) or they get better with experience (Dillon's model). They do this in a way that their proponents hope will help designers and reflect psychological reality.

A suspicion exists that qualitative approaches are inherently vague, are more likely to be rejected by engineers who supposedly like mathematics and, in the light of the aggressive "selling" in the human factors literature of the "hard" quantitative approaches, are somehow less "scientific" however that is measured. This need not be so however. The literature on design (not specifically HCI-related) has clearly demonstrated that designers tend to tackle problems in a manner different from most scientists (Lawson, 1979), relying heavily on heuristics, intuition, and "try it and see" approaches rather than the standard hypothetico-deductive logic based approaches manifest in trained scientists. Qualitative models could well offer the form of guidance more suited to this type of problem solving than any time-consuming but powerful quantitative approach.

As stated earlier, implicit qualitative models abound. All designers and ergonomists, in fact everyone involved in the development of a product, from the marketing department to the specification writers have an implicit model of the users and tasks the end product will support. This model just varies in detail and accuracy depending on one's role. The sales representative presumably has a view of the user as a customer while the marketing person might view users as belonging to certain job, skill and economic categories. These views or representations of the target users are "models" as such.

For present purposes, the main interest is in the models possessed by the designers and ergonomists. The latter participant, by virtue of his probable training in a human science such as psychology, is likely to model the user as an information processor with cognitive dispositions, skills, habits and preferences. This model will probably include detailed knowledge of cognitive components such as short-term memory, long-term memory, mental models etc. and their potential impact on the usability of a computer system. On the basis of task analysis and previous experience, skilled ergonomists can derive a set of user characteristics for input to design specifications. In a very real sense then, this is a form of qualitative modelling.[5]

---

[5] By extension, quantitative model proposers must have their own qualitative models on which they base their formalisms.

Designers, as described earlier, will always implicitly model the user in drawing up specifications. However their models of the user and task tend to be very ill-formed and vague, based on intuition rather than facts. Yet the resultant usability of a product is largely determined at this point by the quality of the implicit model underlying the design and mistakes made at this point are considered to be the most expensive to rectify (Dunn, 1984). The aim of human factors inputs therefore must be to improve this model, either by quantitative or qualitative means. The evidence on balance would suggest that a suitable qualitative model is likely to be more relevant to designers than a formal quantitative one.

What needs to be improved is the explication of these models. Vague descriptions of user characteristics are probably better than nothing but guidelines and handbooks of design principles are rarely successful. Opting to present a more structured view in terms of a framework describing the relevant components of the user-system interaction, as embodied in the present framework, is likely to have more relevance.

Rasmussen (1986) advocates the use of qualitative models in this sense. He argues that quantitative modelling concentrates on one level of behaviour, particularly sensorimotor in well-practiced tasks, which is inappropriate for the type of higher level cognitive functioning of interest to many designers. For him, the major distinction between the two forms of model is not that one is respectable or scientific and the other intrinsically soft and vague but that the qualitative concentrate on broad categories of behaviour while the quantitative focus on specifics. He rejects the traditional engineering argument that the former are merely undeveloped or premature quantitative models and states that designers of computer systems might well find qualitative models of direct relevance to their work in the design of any system where users have some choice on how they will work.[6]

In the case of electronic text design it has been strenuously argued that the quantitative approach is not appropriate. Whether this is a function of current knowledge limitations or inherent failings in the approach is not of direct concern in this thesis, but philosophically at least, the present author's inclinations are in the latter direction. The type of knowledge needed by the designers of the hypertext database at HUSAT was of

---

[6] This is an important but often overlooked difference between current and past technologies. Interactive computer systems afford greater user control than traditional mechanically engineered machines which had to be operated in a set manner. Interesting conclusions could be drawn from this on a range of issues from the changing nature of work to the more enlightened socio-political views of workers in contemporary organisations. Whether such conclusions would be valid however is another issue (see e.g., Eason 1988).

the generic qualitative kind. Furthermore, the act of reading as it is interpreted in the present thesis involves behaviours and cognitions too broad to fit the 10 second boundary of the classic GOMS approach. Therefore, regardless of the ultimate success of a quantitative analysis of cognition, qualitative models do seem to have relevance at this time and are worth pursuing as design aids in HCI.

7.8. The utility of the proposed framework

What use is this framework to designers of electronic text systems? It is possible at this time to specify three potential uses. In the first instance the framework is useful as a guiding principle or type of advance organiser of information (Ausubel 1968) that gives the designer an orientation towards design enabling him to bring relevant knowledge to bear on the problem.

Secondly, by parsing the issues into elements it facilitates identification of the important ones to address. This framework suggest four levels of issue to consider: the user's task and their perception of it; the information model they possess or must acquire; the manipulation facilities they require; and the actual 'eye-on-text' aspects involved.

In the third instance the framework provides a means for ensuring that all issues relevant to the design of electronic text systems are considered. It is not enough that research is carried out on navigation and developers ignore image quality or input devices (and vice-versa). A good electronic text system will address all issues (indeed it is almost a definition of a good electronic text that it does so).

The above applications consider the uses directly to designers at the first stages of system development. In this sense the term designer encompasses any ergonomists or human factors professionals seeking to influence the specification of an application. However the framework also has relevance to later stages of the design process such as evaluation. In such a situation the framework user could assess a system in terms of the four elements and identify potential weaknesses in a design. This would be a typical use for expert evaluation, a common evaluation technique in HCI.

Outside of the specific life cycle of a product, the framework has potential uses by human factors researchers (or professionals less interested in specific design problems) in that it could be used as a basis for studying reader behaviour and performance. The framework is intended to be a synopsis of the relevant issues in the reading process as

identified in the earlier studies. Therefore, it should offer ergonomists or psychologists interested in reader-system interaction a means of interpreting the ever-expanding literature in a reader-relevant light.

Several of these issues will be discussed further in the final chapter of the thesis. However the following chapter outlines supporting evidence for the existence and sufficiency of the four interactive elements of the framework by examining readers' verbal protocols. Then the use of the framework in supporting the derivation of specific models of reader behaviour is empirically examined in a study comparing paper with hypertext.

CHAPTER 8

ASSESSING THE FRAMEWORK IN TERMS OF VALIDITY AND UTILITY

8.1 Introduction

The framework as described in the previous chapter derives from the various analyses on readers' classifications of texts, descriptions of their usage and the experimental investigation of their models for one text type. While it represents an intuitively coherent categorisation of the issues involved in the reading process, it cannot at this stage lay claim to anything more. The obvious questions to ask now are: is this a valid description of reader behaviour? and what purpose does such a framework serve? The present chapter concentrates on answering these questions.

8.1.1 Assessing the Validity of the Framework

In theoretical terms, validity refers to the extent to which any psychological concept or model can be viewed as an accurate representation of that which it purports to describe. It is an important issue for test developers in psychometrics who have devised appropriate means for calculating validity coefficients, i.e., ratings of the extent to which a test really measures what it claims to, be it verbal intelligence, personality or whatever (Anastasi 1990). This is usually achieved by reference to an external objective criterion e.g., scores on a test of aptitude for medicine could be compared with subsequent performance at medical school– a test with high validity should provide scores which correlate significantly with performance at college. For many other psychological constructs such as those typically postulated by cognitive psychologists, e.g., short-term memory buffers, mental models etc., validity is far less amenable to assessment.

In many ways, one can only test the validity of such concepts by trying to prove their invalidity, an odd description perhaps of a process known to scientists as falsification (Popper 1972). That is, the construct (or theory, concept, model etc.) leads to an experimental hypothesis, which the researcher tests by empirical means. If the hypothesis is not supported, the construct or theory is more or less falsified and therefore undergoes modification before the whole process starts again. If the hypothesis is supported, then the theoretical structure is not considered true, but is subjected to further tests. Even if a theory is not falsified after numerous tests it is never considered true, merely adequate, given

current knowledge. This is a somewhat idealised (and Popperian) view of science; in truth, many scientists are loath to subject pet theories to rigid cyclical scrutiny and the scientific process shuffles between contrasting views, numerous modified theories and occasional revolutions (see e.g., Kuhn 1962).

The relevance of these issues to the present case is that if a test of predictions made on the basis of the framework proved positive, this would not imply that the framework was valid psychologically, only that it had not been falsified. The best that can be done is to try and match data from readers to the framework and identify the extent to which the framework would explain these. Any mismatch would be deemed as falsifying the framework and necessitate its modification.

This process is more difficult than it seems however. Obtaining the type of data that would truly test the existence of such concepts as Task Processors or Serial Reading Processors is riddled with philosophical problems, not least of which is what constitutes acceptable evidence? In contemporary cognitive science the development of a working software program that mimics the performance under consideration is seen as the ultimate test (see e.g., Johnson-Laird, 1983). Failure to provide such support is considered a flaw in any proposed model. Thus, according to hardliners, if a hypothesised psychological process cannot be specified as an effective procedure it cannot be considered valid.[1]

This is a strong test for any psychological theory and, not surprisingly, few contemporary theories pass.[2] Such a test could be considered as demonstrating the invalidity of the framework specified in the previous chapter. However, this is not as fatal as it sounds. The framework is not intended to provide a precise model of human mental activity during reading. To test for this would therefore be pointless. In its form as a generic description of the reading process at a level appropriate for design however, it is proposed as valid, and a test of this would be relevant. One test of suitable form would be to examine readers' behaviour and verbal protocols when using a document, parse them into their various components and then relate these to the components in the framework. If the framework is

---

[1] There is a further issue here about the extent to which such a model can be accepted as proof of the process or is just a demonstration of one possible way in which behaviour or output can be caused. This distinction is often labelled "hard" and "soft" respectively in the artificial intelligence community and will not be discussed further here.

[2] It should be noted here that the effective procedure test is not universally accepted by psychologists yet. Furthermore, most psychologists would agree that is is not even appropriate for many levels of psychological enquiry (see Kline (1988) for a strong argument against such testing).

valid, such protocols should provide clear examples of the behavioural and cognitive elements that constitute the framework. If it is an invalid description, the protocols should fail to provide such a match or should indicate the presence of activities not accounted for in the elements of the framework. It is this form of test that will be carried out here.

8.1.2 Assessing the Utility of the Framework

The term utility means relevance, pertinence or usefulness. Accordingly, in this context it implies any appropriate use that may be made of the framework in the electronic text design process. A test of this framework's utility could be made in several ways, ranging from the ideal to the feasible. In this section the potential test scenarios are examined for suitability in the present circumstances.

First, the framework could be presented to designers at the outset of a new design project, under varying degrees of control and manipulation, and its effects on several end products assessed. This would be an ideal test in that it would involve a controlled sample of designers applying the framework in similar task environments. However, under the dual constraints of the commercial pressures in software houses and the limited influence of a single researcher, such a test scenario is unlikely to be feasible. Not only would it require the type of interference in the normal design process that many companies would, quite understandably, actively seek to avoid but the complex analysis that would be required to untangle the dependent variables in such a scenario would be extremely taxing and potentially beyond the abilities of any one researcher.

A near-ideal test might be to present it to a sample of designers working on a range of application development projects and ask for feedback from them on its utility at some specified future date, by which time they would hopefully have had the opportunity to apply it. While this might seem more feasible than the ideal approach, it would still require a commitment to the use of the framework in a commercial enterprise that would seem difficult to justify given its present form and status. Furthermore, it would require both the designers to use the framework as and when they saw fit, and the researcher to rely on sporadic feedback as the sole measure of utility. Given the experiences of many human factors professionals who have handed over design tools in this manner, only to come back later to find them gathering dust on a shelf, and the expected drop out in participation for any data gathering exercise necessitating subjects to initiate their own responses, a high price would almost certainly be paid for a less than perfect study.

With these issues in mind a more suitable test of its utility in the present context would be to apply the framework to an experimental investigation of people reading texts and use it to predict the likely effects of design variables on performance. If the framework is to have any utility to designers it should at least be able to predict gross usage characteristics that are affected by presentation variables. This is a more manageable test of the framework in that it can be carried out without involving a commercial software house and their designers. Furthermore it is the type of test that must be seen as a necessary prerequisite to any real-world testing. It is such a utility test that will be carried out here.

8.1.3. The present studies

In the present chapter, two experimental studies are reported. In the first, an experiment carried out by the author in conjunction with two colleagues at HUSAT offered one suitable test vehicle for the framework. This was an investigation of readers ability to extract relevant information from a text presented in four different formats: one paper and three electronic versions. The data presented here contain part of the original analysis carried out by the design team of which the author was a member (see McKnight *et al.*, 1990a) plus a substantial analysis by the present author alone of data not used by the team in its original work. Though primarily concerned with the validity issue, this experiment also offers some insight into the utility of the framework.

A second experiment designed to test a specific prediction derived from the framework was also carried out solely by the author. This can be seen as a straight utility test in that it represents an attempt to employ the framework to guide predictions of user performance with a text presented on paper or screen. For convenience only (as the descriptors are not precise), the two experiments will be referred to in this chapter as the validity and the utility experiments respectively.

8.2 The validity experiment

8.2.1 Overview

This study examined readers' performance in extracting answers to questions from a short text on the subject of winemaking. The envisaged task scenario was one where an

individual, armed with the document, staffed an enquiry service where people would ask relatively straightforward questions on such topics as the largest winemaking regions of France or the meaning of terms such as 'second fermentation' and so forth. No previous knowledge of wine was required as the answers to all the questions were available in the text.

The aim of the experiment was to examine the extent to which two popular hypertext environments would support such a task compared to paper or a linear electronic text. To this extent it was a very open ended study concerned with exploring the issues rather than manipulating small independent variables. For present purposes the analysis will focus on the concurrent verbal protocols elicited from subjects while also examining some of the performance data.

8.2.2 The application of the framework to the location task

The framework suggests that there are four major components to the task, each of which is represented by a rectangular box in figure 8.1. This schematic representation provides an descriptive model of the likely sequence of events involved in performing the experimental tasks.

Initially, it is suggested, the reader will employ task processing skills to formulate a means of resolving the task. It is probable that for the type of tasks involved in the present experiment, the subject will identify a search criterion from the question and attempt to obtain an answer by finding a relevant match to that criterion in the body of the text. For example, if the question is "What type of wines are produced in the Loire region?" the subject is most likely to select "Loire" as a target and locate references to this in the text until a pertinent section on wine types is located.

Once a satisfactory search criterion is identified, according to the framework, the subject's information model is addressed and used to guide the search for a matching term. It has been shown that the information model of certain text types is well–formed and supports such applications, but in the present context it is expected that the uniqueness of the text would be unlikely to afford a detailed model, particularly at the outset. However, even for unique texts, exposure to them facilitates the development of a map and it is likely that even though a subject lacks an information model at the start, after several tasks he will begin to acquire one. This should be apparent from the verbal protocols.

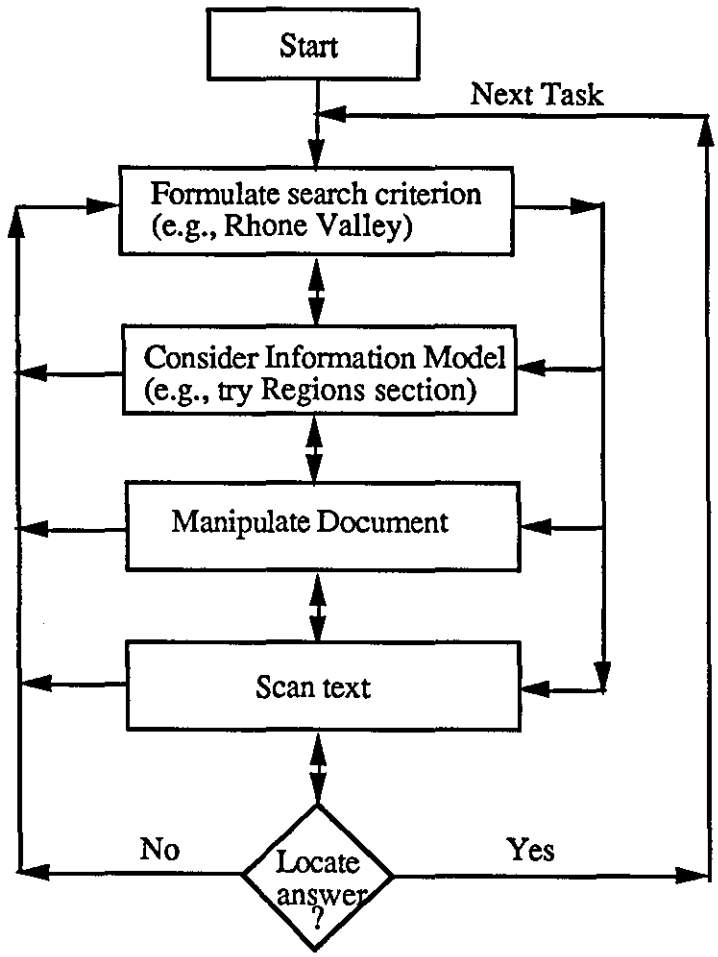


Figure 8.1 A schematic model of readers' behaviour on an information location task.

The subject could, for certain tasks and applications, by-pass or overcome any inherent limitations in his information model by employing the search facilities of the computer. If, at the task processing stage, the reader deduces a specific and infrequently occurring term, the search facilities available on two of the applications could be used to locate the required text directly. Certainly, for the tasks that can be resolved in this way, advantages should be conveyed to subjects using the HyperCard and Word Processor versions. This should be apparent from their performance data.

For other tasks, the manipulations should be less straightforward. For example, one of the tasks required subjects to compare two sections of information before gaining the information necessary to provide an answer. One could imagine the subjects using a paper document opening the first relevant section and keeping a finger on it while searching for the second, prior to flipping between them to obtain the answer. This is a typical reader-paper text interaction but a difficult one to mimic without (and sometimes even with) windowing facilities on screen. Neither of the hypertext applications in this experiment supported windowing of this nature, although this was possible with the word processor version. In such tasks, one would hypothesise advantages to paper over electronic text. The advantages and disadvantages of manipulating text should be manifest in the subjects' protocols.

Once the reader has searched and manipulated the text, the framework suggests that a scanning type of reading follows. In other words, it is not expected that subjects will read large amounts of text in a serial fashion while performing these tasks but will jump and skim read sections looking for cues and target words. From the work on proofreading text on paper and screen reviewed earlier, it is clear that in general, the advantages lie with paper. Although few researchers have examined scanning as opposed to proofreading, there is no reason to assume that a similar advantage to paper does not hold for this type of reading style too.

If the target is successfully located at this point then the task is completed and the subject can start on the next one. This initiates a sequence of events similar to those just outlined, though with each subsequent completion it is expected that the information model becomes more elaborate (i.e, a mental map of the document is being formed) and familiarity with the requisite manipulation skills grows. This should also be apparent in the protocols.

8.3. Method

8.3.1 Subjects

16 subjects participated in the study, nine male and seven female, age range 21–36. All were members of HUSAT staff and all had experience of using a variety of computer systems and applications.

8.3.2 Materials

The text type under consideration was a document titled "Introduction to Wines" by Buie and Hassell (1982), a basic guide to the history, production and appreciation of wine. This document was widely distributed in hypertext form by Ben Shneiderman as a demonstration of the TIES (The Interactive Encyclopedia System) package (now known as HyperTIES). In the TIES version, each topic was held as a separate file, resulting in 40 individual small files. For the Hypercard version, a topic card was created for each corresponding TIES file.

In its paper format this text consists of 13 A4 pages of text with no figures and would thus be most aptly described as a booklet or essay type text. In order to place a structure on the document that would facilitate its presentation as a paper text the topics were retained in the linked groups of the hypertext original but ordered from start to finish in a manner that seemed intuitively sensible to the experimenters. Thus an introduction was followed by a general overview of the processes involved in manufacturing wine before specific countries and regions were presented. This structure was retained faithfully for the word processor version.

In order to test this intuitive arrangement for suitability a quick pilot test was carried out by the author. This involved asking subjects to order a set of cards, each of which had a term on it referring to the title of each of the files in the TIES version. These were wine related terms such as "Bordeaux", "Production" or "Aperitifs" rather like a list of contents. Three subjects were each asked to group these into what they perceived to be a suitable single-document structure. The results confirmed the structure of the experimenters i.e., groups were formed out of countries and subordinate regions, wine manufacture, and particular wines and grapes.

The Hypercard and word processor versions were displayed on a monochrome Macintosh II screen and the TIES version was displayed on an IBM PC colour screen. The paper version was a card-covered spiral bound, A4 text.

8.3.3 Task

Subjects were required to use the text to answer a set of 12 questions. These were specially developed to ensure that a range of information retrieval strategies were employed to

answer them and that the questions did not unduly favour any one medium. The answers to all questions were specifically mentioned in the text.

8.3.4 Design

A four-condition, independent subjects design was employed with presentation format (HyperCard, TIES, Paper and Word Processor) as the independent variable. The dependent variables were speed, accuracy, access strategy, subjects' estimate of document size and verbal protocols.

8.3.5 Procedure

Subjects were tested individually in the usability laboratory at HUSAT. This consists of a simulated office environment containing video cameras and sound recording equipment, separated from an observation/control room by a door and one-way mirrored window. The experimenter described the nature of the investigation and introduced the subject to the text and system. Any questions the subject had were answered before a three minute familiarisation period commenced, during which the subjects were encouraged to browse through the text. After three minutes the subjects were asked several questions pertaining to estimated document size and range of contents viewed. They were then given the question set and asked to attempt all questions in the presented order. Subjects were encouraged to verbalise their thoughts and a small tie-pin microphone was used to record their comments. Movement through the text was captured by video camera situated non-intrusively directly behind them.

8.3.6 Experimental Hypotheses

As the model of likely behaviour deduced from the framework suggests, a simple prediction about the most suitable application for these tasks is not possible. Paper would seem to have certain advantages over all electronic versions in some circumstances while it is possible to see advantages for certain electronic versions in others. However, several experimental hypotheses suggest themselves on the basis of the framework:

1. The size of the document and the lack of any specific superstructural model of the information space should convey a general advantage to paper and word processor versions in the first instance. This should manifest itself in problems

estimating the document size and greater navigational difficulties in the hypertext conditions.

2. The access mechanisms in the electronic versions should convey advantages to them over paper for certain tasks e.g., those supportable by using search facilities. Therefore the HyperCard and Word Processor versions should lead to faster completion rates on those tasks than the HyperTIES and Paper versions.

8.4. Results[3]

8.4.1 Estimating Document Size

The results of the study generally support part of the first hypothesis, hypertext users had difficulty assessing the document size accurately while subjects in the linear conditions were far more accurate. After familiarisation with the text, subjects were asked to estimate the size of the document in pages or screens. The linear formats contained 13 pages, the Hypercard version contained 53 cards, and the TIES version contained 78 screens. Therefore raw scores were converted to percentages. The responses are presented in Table 8.1 (where a correct response is 100, scores above and below this number reflect over- and underestimates respectively).

Condition	TIES	Paper	HyperCard	W.Processor
Subject				
1	641.03	76.92	150.94	92.31
2	58.97	92.31	56.6	76.92
3	51.28	76.92	465.17	100.0
4	153.84	153.85	75.47	93.21
Mean	226.28	100.0	187.05	90.61
SD	280.41	36.63	189.84	9.75

Table 8.1 Subjects' estimates of document size.

[3] Some of the data presented here are used by kind permission of my co-workers Cliff McKnight and John Richardson. However, interpretation of this data in relation to the framework is the sole responsibility of the present author and does not necessarily reflect the views of either of these researchers.

Subjects in the linear format conditions estimated the size of the document reasonably accurately. However, subjects who read the hypertexts were less accurate, several of them over-estimating the size by a very high margin. While a one-way ANOVA revealed no significant effect ($F_{[3,12]} = 0.61$, NS) these data are interesting. They suggest that subjective assessment of text size as a function of format is an issue worthy of further investigation and thereby confirm the importance of this issue as indicated in the information model component of the framework.

8.4.2 Navigation

The other part of the first hypothesis was that navigation would pose more difficulties for users of the hypertexts than for the others. As stated in the literature review, a general measure of navigation is non-existent but relies on the interpretation and operationalisation of the concept by individual researchers. For present purposes it was assessed by examining the proportion of time spent viewing the Contents/Index (where applicable) by each subject as a percentage of total time. This provided a highly objective behavioural measure rather than any indication of subjective difficulty. These scores are presented in Table 8.2.

Condition	TIES	Paper	HyperCard	W.Processor
Subject				
1	53.28	2.72	47.16	6.34
2	25.36	1.49	19.1	13.93
3	49.5	10.24	17.5	12.87
4	30.84	5.36	23.4	7.54
Mean	39.74	4.95	26.79	10.17
SD	13.72	3.88	13.81	3.79

Table 8.2 Time spent viewing Contents/Index as a percentage of total time.

This table demonstrates a very large difference between both hypertext formats and the linear formats. A one-way ANOVA revealed a significant effect for condition ($F_{[3,12]} = 9.95$, $p < 0.005$). Even using a more rigourous basis for rejection of the null hypothesis in *post hoc* tests than the 5 per cent level, i.e., the 10/k(k-1) level, where k is the number of groups, suggested by Ferguson (1959), which results in a critical rejection level of $p < 0.0083$ in this instance, post-hoc tests revealed significant differences between Paper and TIES ($t = 4.90$, d.f. = 6, $p < 0.003$), between Word Processor and TIES ($t = 4.16$, d.f. = 6, $p<0.006$) and between Hypercard and paper ($t = 3.06$, d.f. = 6, $p < 0.03$). Thus,

interacting with a hypertext document may necessitate heavy usage of browsers or indices in order to navigate effectively through the information space.As a result of this analyses in conjunction with the estimates of document size it seems reasonable to reject the first null hypothesis.

8.4.3 Searching for precise information[4]

It is expected that when subjects seek information for which they can formulate accurate search terms, applications that offer such facilities should lead to faster and/or more accurate task completion rates than those which do not. However, use of the search facilities rests on several factors: the realisation of their presence; the willingness to use them; and the ability to use them correctly.

It was expected that the users in the present sample would be familiar with search facilities and realise that they existed in the relevant applications. The willingness to use them is more difficult to predict and it is possible that users will only employ them when other tactics fail. Using them correctly is a skill and any realistic model of user behaviour must allow for possible errors.

In this study, three of the tasks were supported by the search facilities in the HyperCard and WORD conditions. The mean time per subject on these tasks is shown in Table 8.3.

	No Search Facilities		Search Facilities	
Condition	TIES	Paper	HyperCard	W.Processor
Subject				
1	194.33	66	112	97
2	79.67	34.67	44.67	59.33
3	281.67	233	39.33	81.33
4	122.67	170.67	36.33	76.33
Mean	169.59	126.08	58.08	78.50
SD	88.43	91.99	36.11	15.57

Table 8.3 Mean times to perform tasks supported by search facilities

Analysis of time taken to locate information using the various applications confirms the

[4] The analysis in this section is solely the work of the present author

view that those applications supporting search facilities would be more effective for such tasks. A one-way ANOVA comparing those applications with search facilities and those without showed a significant effect in the hypothesised direction ($F[1,15] = 6.10$, $p < 0.03$) thereby allowing therejection of null hypothesis two. Interestingly, not all subjects used the search facilities for all possible tasks, as suggested above.

8.5 Evidence for the interactive elements from readers' protocols.[5]

The data described above demonstrate that the framework can be employed to guide reasonably accurate predictions about reader task performance. However, for present purposes, the main aim of the study was to examine the extent to which the framework can be seen as a valid representation of the issues involved in reader-text interaction. To test for this, the verbal protocols of the subjects were examined.

Each protocol was transcribed from a video tape according to a predefined classification scheme derived by the present author in conjunction with the other members of the research team. This scheme captured the verbal utterances, the time they occurred, the actions performed by the subjects and any further behaviours deemed relevant by the transcriber such as subjects having difficulties with an application or making an error in their answer. These were subsequently examined by the author in order to identify verbalisations that mapped onto the framework. Protocol data are rich and complex therefore not easily reduced to a simple presentable form. In the present context isolated sections from a selection of subjects in a manner akin to Suchman (1988) seemed to be the most appropriate means of highlighting the existence of the elements of the reading process described in the framework. Accordingly several examples from different subjects are presented to provide an insight into the reading process from the point of view of the reader. However, in order to counter the argument that the selected examples may be merely the best selection of quotes rather than typical examples of readers' verbalisations, a full protocol of a subject performing all tasks is included as Appendix B.

Example 1 describes a typical section of protocol from one subject. The protocol presentation format involves a time scale in seconds, a transcription of the user's verbal protocol and a description of user action on the system.

---

[5] The analysis in this section is solely the work of the present author

TIME	COMMENT	ACTION
00:00	O.K. I'm going to the Index to see if any of these terms are mentioned....	Selects INDEX Button
00:03	Don't appear to.....	Reads Index
00:25		Selects CONTENTS button
00:35	I'll have to look in..... probably section 2, The Making of Wine.....	Reads Contents
00:58	I'll go to sweetness because it's the only term in the contents list that really.....refers to taste..	Selects SWEETNESS

This represents a short section (1 minute) of a subject looking for specific information in a HyperCard stack. While on the surface it might seem trivial, it is obvious that much information processing is occurring in the subject's mind. Firstly, without hesitation, he has accessed the Index. This is logical behaviour as Indices provide references to and the locations of material contained in the text. However this implies that not only does the subject have expectations of what he will find in that section of the text but that he has formulated a means of task resolution and acted upon it rapidly. Without some form of mental representation of the order of the text, the manner in which it might be structured and the access mechanisms available within it i.e., a model of the information space and the manipulation facilities available, such rapid and meaningful processing would not be possible.

It would be virtually impossible to determine the order in which the cognitive processing went for such an interaction. Obviously he knew his task and had a formative map of the document (at this stage the reader had already spent three minutes familiarising himself with the text prior to commencing the trial). These elements must have interacted but whether that interaction was of the form TP—>IM or IM—>TP is probably not important from a designers point of view. In reality it was probably a case of cyclical interaction between both. What is certain is that the subject decided on a course of action, manipulated the

information space and rejected the first attempt at problem resolution in a matter of seconds.

In fact, in the first few seconds of this example it is possible to see all four elements of the framework in action. Once he decides to go to the Index the Manipulation element operates (button selection) and the Reading Processor scans down the list of topics to reveal no mention of the target item.

The first interaction of another subject, this time a subject in the word processor condition, is presented below.

TIME	COMMENT	ACTION
00:10	Right, I'm going to go to the Contents and try and get some idea of where this might be.....	Drags SCROLL bar up
00:24	I think it's probably in "The Making of Wine"	Still reading CONTENTS
00:38	Oh is it Sweetness and Body?	
00:42	I'll just go there and check if they're the two.....	Scrolls down the text first then clicks in SCROLL bar until she reaches relevant section. Reads them.
00:58	So the two things are..... Sweetness and Body.....	

Again the interaction of several sub-processes of the reading task can be identified. Upon reading the question, the subject's first decision is to go to a section of the text that might offer guidance pertinent to the task. As before the expectation that Contents sections adequately map the information available in the text and the fact that they are located in a particular place should be noted. At this point an interaction between TP and IM has occurred followed by a quick burst of MLS and SRP activity to move up the text on screen and identify the relevant section as the desired one.

Upon reading the Contents (SRP activity) the subject identifies two sub-section headings that seem relevant to her representation of the task (TP activity) and decides to go to that

section (IM activity) to check if they are relevant. At this point she scrolls down the page before deciding that the scroll bar would be a better option (MLS activity interacting with TP activity and IM activity - the subject must decide at some point that the information is located suitably far away to justify a faster or more efficient scrolling method). Upon arriving at relevant section (rapid SRP, IM and MLS activity) she ceases scrolling and starts to read the text until she confirms her opinions.

At this stage in both examples only the crudest information model of the text exists. Since this is a unique text type to which the subjects would not have been previously exposed there would be no existing structural models for them to employ and only maps could be formed. With increased exposure this map is elaborated allowing more accurate predictions of what is located where and the type of information available within it. Thus one observes numerous comments to the effect that "I've seen this before somewhere" or "I've an idea where this one is..." and so forth. The following example is from a subject using the word processor document who is now on his fifth task.

Here we see the emergence of landmark knowledge of the information space (I've seen nothing on that.....I've just passed a section on... etc.) and knowledge that can allow the reader to make informed judgements about locations that enable him to evaluate (and in this case, reject) hypothetical locations of information (Grapes on page 1?.....No.). Such processing is only feasible where the individual has at least a rudimentary map of where things are in the document, what they are next to, whether or not he has seen them before and what type of information a section may contain. As suggested by the framework this type of knowledge is likely to be picked up by subjects as they become familiar with a document.

Also pertinent here are the limitations of the manipulation facilities for scrolling text in word processors. In going to the Index in the first instance the user drags the scroll bar down to the bottom of the window with the result that he overshoots the start of the index section and needs to scroll gently back up (a total activity taking approximately 10 seconds). This is similar to, but probably more awkward than using indices at the back of books i.e., overshooting to the back of the book so that you need to page back to the index section is a common experience during reading but shouldn't waste 10 seconds. Technology should make such a process simpler (or unnecessary), not more difficult. This is an action that is well supported in the "point-and-click" facilities common to the hypertext systems.

TIME	COMMENT	ACTION
06:07	To the Index then.... I haven't seen anything on this before....	Drags SCROLL bar down
06:09		Scrolls up to top of Index
06:17	Grapes on page 1? No.....	Reading Index " " "
06:25		Drags SCROLL bar up to Page 2
06:30		Scrolling up to top of section Reads section
06:42		Scrolls down and reads next section
06:56		Scrolls up and re-reads section
07:00	It must be in the body of the report then.....	
7:30		Spends 30 seconds scrolling down and reading the following four sub-sections
07:33	O.K. ..dessert wines..	Locates answer
07:42		Reads next question
07:45	I've just passed a section on aging...	Scrolls up to that section.

These examples are typical of the protocols elicited in this study. The framework proposed here seems to be supported by the evidence from protocols of readers using both paper and hypertext versions of a document in the following ways:

(i) There is evidence of the existence of each of the elements i.e., subjects verbalise thoughts which confirm attention changing from attributes of task, information model, manipulation and straight reading of the text.

(ii) There is evidence of the interaction of the elements in both serial and non-serial fashions (i.e., interactions do not necessarily follow the strictly linear sequence: TP—>IM—>MSF—>SRP but combine in sequences which reflect the reactive nature of the reading process).

(iii) There is no evidence from any of the protocols to suggest other elements need to be incorporated, all relevant verbalisations and behaviours are classifiable as belonging to one of these categories (though of course, these categories or elements are rather general and hide a myriad of complex cognitive issues as stated earlier).[6]

It seems as if the framework provides a relatively parsimonious account of the types of utterances solicited from subjects performing routine tasks with a text. No attempt has been made here to compare the verbal protocols between conditions to identify any differences that might exist between them. For example, it is possible that subjects in the hypertext condition manifested less comments on their information model than subjects in the paper condition, or most comments in all conditions were about manipulation rather than scanning issues. While such findings might be of some interest in terms of the psychology of reading they are secondary in importance to the aims of the present work i.e., the development of a useful descriptive framework of the reading process to aid electronic text designers.

A further reason for not pursuing a more quantitative analysis of the protocol data is that such a procedure is difficult to perform objectively. In the examples described above one can identify the general sequence of activities and their relationship to the framework. However, to force every protocol into an all-or-nothing categorisation whereby each utterance is classified as belonging to one or other element in the framework would hardly be informative and would lead to a mass of numeric data that added little or nothing to the present description of the framework. Many of the interesting utterances concern the rapid interaction of several elements or the continual interchange between two non-adjacent elements. If the quantitative categorisation was to take account of all elements, the possible interactions between elements and the timeline associated with each utterance and try to relate these to each of the four conditions in the experiment it is likely that the resultant analysis would over-complicate the data to the point of meaninglessness.

In summary, the framework posits the existence of four elements of concern to the reader of any text. The verbal protocols support the existence of these and suggest that the interactions between these elements is of the general form described in chapter seven. The

---

[6] Not surprisingly there were verbalisations that were deemed irrelevant, such as comments to the experimenter regarding the time, the ease or difficulty of particular tasks or quips about the study. However these amounted to a very small proportion of the total data elicited and can safely be considered inconsequential with regard to the framework.

framework provides an adequate account of the type of processes carried out by readers of both electronic and paper texts.

8.6 The Utility Experiment[7]

8.6.1 Overview

The previous study also showed that the framework can support accurate predictions about reader performance with a text as a function of presentation medium. The present experiment extends this to another text type and employs a further hypertext application (GUIDE) not used in the last experiment.

As stated in chapter four, a hypertext journal article is unlikely to be a complete replacement for the paper version. On the basis of the task analysis and literature review it would seem that for straight reading of the text, paper would be preferred and be more usable. However, for the other forms of use to which such texts are put hypertext is likely to offer certain benefits. This text type would therefore seem a useful test of the utility of the framework.

8.6.2 Applying the framework to the description of academic article usage

A journal reader approaches a text with a task or set of tasks that he hopes to resolve, no matter how ill-specified. According to the framework it is suggested that the readers apply their model of the text structure to the task in order to direct their activities. Thus they decide if they need to look at some part of the article more than the other, where that part is located, where the other relevant articles might be and so forth. If the electronic version maintains the paper structure there should be no differences between the media at this stage i.e., their well-developed model would be equally relevant to either medium.

Readers then manipulate the text and locate the section(s) relevant to their needs. Traditionally this would have been difficult with electronic text but the availability of hypertext applications eases the manipulation task considerably, particularly where text is broken or "noded" into selectable chunks. In the case of the article, jumping to various sections and headings should be facilitated on screen, though location of particular blocks

---

[7] This experiment is solely the work of the present author

of text within a larger body of text is unlikely to be so easy.

Once at the relevant section it is probable that readers adopt one of two reading styles: straight serial reading from the start or quick scanning. In reality readers probably adopt a mixture of both. Where the reader adopts a serial reading style, paper is likely to be better than hypertext. This seems probable given the weight of evidence showing a performance deficit for proofreading speed from screens and the difficulties readers have with lengthy electronic texts. However, the differences between the two media are likely to be lessened where the amount of text to be read is small (for argument's sake let us say a screenful). Where the reader is scanning the material and it is not lengthy, there is likely to be little difference between the media (assuming image quality of the screen is good). If the text is broken into various small sub-sections within a section and the reader has an idea where he wants to go, hypertext should convey advantages over paper.

Accordingly a simple descriptive model of user behaviour for a particular task, for example, checking a detail in the method section such as the number of subjects employed or the type of equipment used could be derived. In circumstances where the paper article's structure is retained in hypertext we would assume no difference between the media until the MLS and SRP elements of the framework are invoked. At this point one would expect an advantage to the hypertext version for getting to a headed section but an advantage to the paper version for the scanning or serial reading phases of performance. A descriptive model of such a task is represented in figure 8.2.

According to this model the task basically consists of quick target identification, a rapid application of the IM, and then a sequence of MLS <—>SRP interactions. The latter interactions dominate the task according to the framework though each element is not used in equal proportion. Given the task involves scanning text in a specified (i.e., given) area there is likely to be a bias towards SRP involvement. MLS activity should be rapid while SRP activity could be extended.

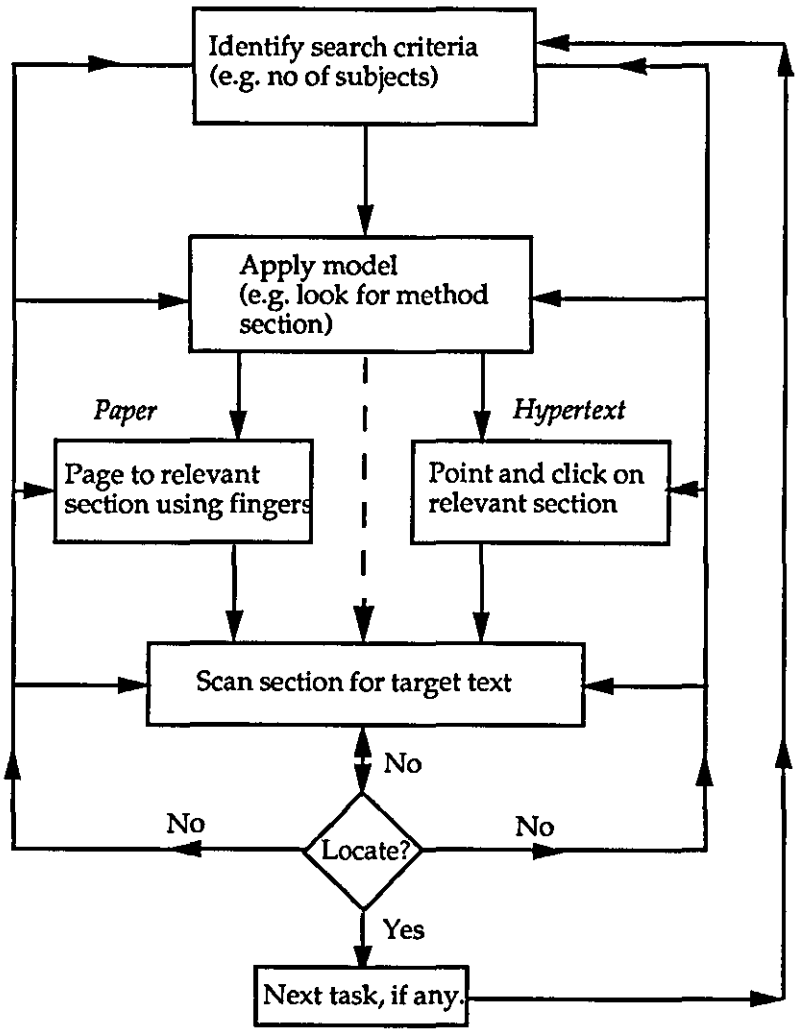


Fig. 8.2 A schematic model of readers' behaviour for journal task.

From what is known about reading from screens it is clear that for SRP activity, paper should be faster than hypertext (approximately 20% faster according to current estimates). While manipulation can prove problematic for electronic texts, the "point and click" approach of GUIDE is familiar to the subject sample employed here and where the targets can be directly addressed from one screen, advantages to hypertext should ensue. However, given the relative proportions of time estimated to be spent in either activity, this should not be enough to offset the reading speed advantage to paper.

Though there should be an overall advantage to paper it is possible that for targets not requiring large SRP activity the ease of manipulation with GUIDE might prove sufficient to

give hypertext an advantage. This would occur in situations where the target sentence was situated at the start of an opened section. Conversely, the speed advantage to paper should maximise the differences between the media for targets situated towards the end of lengthy section. These differences suggested by the model are tested in this study.

8.7 Method

8.7.1 Subjects

Twelve subjects (age range 22-35, mean age 27; six male/six female) participated voluntarily in this study. All were professional researchers experienced in the use of academic journals and frequent users of personal computers.

8.7.2 Stimulus materials

Two academic articles were selected according to the criteria of similarity in terms of length, broad subject matter (on computer human factors) and conformity to the general superstructure of articles described in chapter six. Good quality photocopies were made for one condition and GUIDE versions created for the other. The hypertext versions were presented, black on white on an Apple Macintosh II. Screen Recorder™ was used to record (non-intrusively) subjects' performance with the hypertext.

8.7.3 Task

Subjects were required to locate 32 sentences in the academic articles. These were divided into four task blocks of eight sentences each so that each subject located sentences in both texts using both media (in order to control any possible text biases or presentation order effects). The sentences were presented on stimulus cards which stated in which section of the text (Introduction, Method, Results, Discussion) the sentence was to be found. The task was designed to be a simulation of the situation common to readers of these texts which is checking a detail of the paper when they have a fairly reliable notion of where the target sentence is located but are not likely or able to use the search facilities (i.e., they cannot formulate an appropriate search string).

The sentences were selected so that equal numbers came from all four sections, they were of approximately similar length (i.e., less than two printed lines) and did not contain eye-

catching words (e.g., all-capitals) or symbols (e.g., numbers). The hypertext versions of the texts were made so that line lengths were comparable to the paper ones thereby ensuring a similar typographical form for both media. As well as being situated in particular sections, the target sentences were further distinguishable in terms of their within–section location. Thus in all but the Method section for which it would be impossible, the sentences in each section were selected so that they were at the start or towards the end of the relevant section. The qualification for being located at the start was determined by the presence of the sentence in the first full screen of text that was presented upon opening a section in the hypertext. This allowed further analysis of subject's performance i.e., the effect of scanning large amounts of text on either medium.

8.7.4 Design

A two condition (paper x screen) repeated measures design was employed using two texts. All subjects performed the task twice on paper (once per text) and twice with hypertext (once per text) with order of texts and presentation medium counterbalanced to avoid any systematic ordering effects. The independent variable was presentation medium and dependent variable was speed of task performance.

8.7.5 Procedure

Subjects performed the experiment in an experimental room at HUSAT. The computer was placed on a desk which was free of other materials allowing them to perform both the paper and the hypertext tasks without changing desks. The experimenter sat at the edge of the desk in a position where he was able to see the screen and the document being read by the subject at all times.

All subjects were introduced to the concept of hypertext and the specific workings of the GUIDE package. Most expressed familiarity with the concept but only three subjects had actually used GUIDE. Subjects were then encouraged to interact with the application until they were comfortable with it, at which point they performed five trial tasks to consolidate their training. If they still experienced any difficulties further blocks of trial tasks were available. However, no subject asked for or appeared to require extra training.

Subjects were informed that they would be timed for each individual task. Timing started from the moment the experimenter handed them the stimulus card containing the target

sentence until the time when they successfully located it. Successful location was marked by a verbal statement of the fact and pointing to the sentence with finger or cursor, enabling the experimenter to confirm that the target was successfully located. Upon location, the experimenter noted the time lapsed and ensured that the subject closed the paper version or went to the top-level of the GUIDE document before commencing the next task. Time was recorded in seconds using a stopwatch. There were two minute rest periods between each block of trials.

At the end of the experiment subjects were asked to describe their general ratings of the hypertext version and its suitability for journal article presentation.

8.7.6 Experimental Hypotheses

Given the task and the conditions under which they are presented it was expected that there would be only two levels of difference between the media: the manipulation and the skim reading ones (MLS and SRP elements) as outlined above. Given the variation in locations and the estimated proportion of time spent on each activity, three experimental hypotheses were proposed:

1. There would be a significant difference overall between the two presentation media for the completion rate of tasks with paper proving faster than hypertext.

2. Subjects should locate information in lengthy sections of the texts faster with paper than with hypertext.

3. Subject should locate information in short sections of text faster with hypertext than with paper.

8.8 Results

8.8.1 The Effects of Medium, Text and Question on Performance

A three-way, 2 x 2 x 8 ANOVA (medium by text by question) with repeated measures on all factors was carried out on the data using the MacSS statistics package on an Apple Macintosh Plus. Although the texts were selected for similarity it was decided to test for

text in case it was producing an effect that was not dependent on the variables controlled for in their selection such as idiosyncratic writing style or vocabulary. The output from this analysis package is summarised in Table 8.4.

Source	df	SSQ	MSQ	*f*	p
Guide/Paper (A)	1	33189.85	33189.85	14.909	0.0029
Text 1/2 (B)	1	2470.51	2470.51	1.7418	0.2119
AB	1	133.01	133.01	0.2187	0.6524
Question (C)	7	122309.5	17472.79	12.655	0.0001
AC	7	29399.67	4199.952	2.7579	0.0131
BC	7	16546.74	2363.82	1.3393	0.2429
ABC	7	20704.99	2957.856	1.43105	0.2045
Error	352	565404.37	1601.712		
Total	383	790158.64			

Table 8.4 ANOVA summary table for utility experiment

These results indicate a significant effect for medium, question and the interaction between medium and question. There was no significant three-way interaction effect or significant effect for text type as expected.

Clearly, there was a significant effect for medium with paper proving to be faster than hypertext for this set of tasks. Thus null hypothesis 1 can be rejected. Mean time per task with hypertext was 52 seconds compared to 33.5 for paper (i.e., paper was approximately 35% faster than hypertext which is just outside the range of speed differences between the media typically reported in the literature).

The significant effect for question was also expected. Searching for targets in text sections of varying length should lead to speed differences with shorter sections affording faster location. This requires no further explanation. Mean times for each location confirmed the direction of the differences, e.g., location times for questions 3 and 4 (Method section) were 14.65 and 16.46 seconds respectively, while mean location times for questions 1 and 2 (Introduction section), were 40.77 and 56.60 seconds respectively.

8.8.2 The effect of target position on performance

According to the framework, subjects should have been able to locate sentences that

occurred in the short sections of the texts (e.g., in the Procedure sub-section of the Method section) faster in hypertext than on paper. The opposite should hold for sentences embedded in longer text sections such as the discussion (though where sentences occur at the start of such sections hypertext should regain the advantage). Thus the advantages of either medium should be tested between questions.

The significant effect for question, and more importantly the significant interaction effect for medium and question give an indication of what happened in the present case. By examining the mean times per task for each medium it is clear that the advantage to paper is most obvious for tasks involving the scanning of large sections of text. No such difference holds in when the target sentence is located in shorter sections. The unweighted marginal means for task by medium are presented in table 8.5.

Target	Hypertext	Paper	p
Introduction/early	46	35.5	ns
Introduction/late	63	50.5	ns
Method 1	15	14	ns
Method 2	17	16	ns
Results /early	75	39	.05
Results/late	93	42	.05
Discussion/early	50	46	ns
Discussion/late	57	24	.05

Table 8.5 Mean times (seconds) per question for each medium

P values were obtained by using the marginal means to calculate a value for t according to the formula $t=(mean1 - mean2)/\sqrt{(mse/n1 + mse/n2)}$ (as described by Ferguson 1959, p. 238), where mse=1522.86 and df= 77 (i.e., 1 x 7 x 11).

These data provide a better view of the results than the overall difference between media. Far from being enormously better than hypertext for this type of task it can now be seen that the advantage to paper is maximised mainly for location of material that is situated towards the end of lengthy sections. Although this difference was non-significant for target sentences in the Introduction, it was still large and in the hypothesised direction. This supports experimental hypothesis 2 for which the null hypothesis can now be rejected. For targets that occurred in the first few paragraphs of a section or in sections that did not contain large expanses of straight text (e.g., the Method sections) then subjects performed as well with hypertext as with paper. However they did not perform significantly better, as predicted, therefore necessitating the retention of null hypothesis 3.

The significant effect for targets in the early parts of the results section runs counter to the suggestion drawn from the framework. Examining these tasks on the Screen Recorder output confirmed what had been suspected by experimenter observation, i.e., subjects regularly missed the target on first exposure to the section and read serially through to the end before returning to the relevant part of the text. In all, eight subjects failed initially at least once to locate the target even when it was first on screen. Of these, six failed to do so on more than one occasion and two subjects missed the same target on two occasions (i.e, they re-read a section more than twice before locating the target). Though there were no equivalent records of subjects in the paper conditions, the experimenter noted only one subject doing this.

8.9 Discussion

The present investigation was intended to simulate the type of task performed by readers when searching for a specific piece of information in a familiar text type. In the academic article situation this would involve searching for a reference, checking a detail in the design, or finding the major results etc. It was hypothesised that this would be the type of reading task for which hypertext would in part, offer suitable, perhaps even advantageous support.

The task consisted primarily of two components of the reading process, manipulation and skimming of texts, which were related to two elements of the framework. Hypertext offered clear speed advantages for getting to a section, it was merely a matter of pointing and going. Though not timed at this level, paper required the use of both hands to flick through the pages and subjects often paged past the sections they wanted in their attempt to jump directly to it. This supports the original view of hypertext as a potentially advantageous medium for manipulating large texts.

However, once at a section, readers performed faster with paper, particularly when the target sentence was not immediately obvious. This is explicable, at least in part, in terms of the image quality hypothesis discussed earlier. The statistically significant advantage to paper overall predictably emerged as a result of concatenating all the actions into one performance score: total time. Since the percentage time spent reading was normally greater than that spent manipulating the text this weighted the final measure in favour of the task

component known to be best supported by paper.

These results support two out of three of the hypotheses derived from the framework. However, the predicted advantage to hypertext for locating text in short sections of text never materialised. In fact, the general trend of the results has been to distort the predicted differences further in the direction favouring paper than was expected i.e., the reading speed difference was larger than normally observed. How can this be explained?

Several issues to do with general presentation on screen are worth considering here. Although the hypertext was presented as black on white, the font (New York 12 point) was chosen to retain similarity to the printed paper font rather than optimising the screen display. On the basis of this study and that of the experiment on readers' models reported in chapter six, this was a mistake. Paper fonts are optimised for paper reading, screen fonts should be optimised likewise.

One subject remarked that Geneva would have been a more suitable font and it was clear from talking to subjects afterwards that text presentation is generally viewed as poorer on screen than on paper regardless of font. A variety of associated reasons such as angle of viewing, flicker, and other subjective measures which have failed to produce clear explanations of empirical differences in the work of Gould *et al.* (1987a,1987b) *inter alia*, are still reported by subjects as disruptive.

The value of trying to mimic the line length of the paper article on screen was also questioned by two subjects. They both felt that for screen reading, a wider display with increased inter-line spacing would have been more readable and helped them to identify relevant target characteristics more easily. There is some evidence in the literature to support these suggestions (e.g., Kolers *et al.*, 1981, Duchnicky and Kolers, 1983).

In terms of the descriptive model derived from the framework there are obviously few, if any, modifications required to explain these findings. It led to the accurate prediction that paper would be better than hypertext both overall and for locating sentences situated in the later parts of lengthy sections. The shortcoming of the other hypothesis seems to result not from shortcomings of the framework or any of its postulated elements but from the task and sensitivity of the measures employed.

The sentence location task was intended to simulate the type of reading scenario where a

person searches relatively familiar material to identify a certain detail. Behaviourally at least the experimental task employed here matched this. However, cognitively it is difficult to maintain the comparison. Subjects in the present situation reported trying to locate the required sentence in a simple pattern match fashion, i.e., they focussed on a key word or phrase in the target sentence and searched the text closely for anything that matched this without considering the meaning of the material being attended to. In the real reading situation one would expect the reader to be more influenced by the context of the material and seek to relate the content of the currently attended-to paragraph to his information needs.

In this way, real reading would involve the narrowing of the search space according to the context of the target's location e.g., if a reader wants to find a sentence about difficulties with the experimental procedure he is likely to appreciate the relevance of other words or sentences which mention other problems or procedural issues. There was no evidence of subjects in the present study actually trying to relate the content of the target sentence to the experimental text beyond the cues they were given for searching in a specific section. Indeed, the typical reading style manifested by subjects was a straight serial read from start to target (or finish) of the prompted section, scanning every intervening word. In reality, one would expect to see readers jumping about within sections, ignoring paragraphs which the first sentence indicated were unlikely to contain the required details. This would have the effect of speeding up this process and thereby lessening the proportion of task time spent at the SRP level of the framework with commensurate benefit for hypertext users.

A further source of potential bias in the study was the assumed lack of training required by subjects in using GUIDE. Although all subjects expressed confidence after the familiarisation period with the use of this application for manipulating the text and there were no instances of subjects being unable to open or close sections of the hypertext during task performance, most reported afterwards that they would require a lot more use of GUIDE before feeling as comfortable with it as they were with the paper texts. Several subjects certainly manifested non-optimum use during task performance e.g., opening irrelevant sub-sections and failing to close them before opening another which resulted in large text sections to scroll through if they went back to the "start" to re-read a section.

The design of the experiment, with its demand that subjects complete all tasks failed to allow for any potential speed/accuracy trade off. Since many of the subjects overlooked the target on initial exposure and were forced to re-read sections, sometimes more than once,

their performance (i.e., speed) scores deteriorated rapidly. Had subjects been given the option of giving up, the speed differences might not have been so large. Furthermore, for most electronic texts, when faced with a situation where a target is proving evasive, the search facilities are likely to be used. These were not supported in the present task as it was an attempt to simulate the type of interaction where the reader has only an ill-formed idea of the specific details of the search, therefore being more likely to use recognition over recall to aid location. Future work would do well to consider such task effects and select accordingly.

In summary, the data suggest that hypertext can be as good as paper for information location tasks when the reader has (or is given) an accurate information model and is not required to read lengthy sections of text. Where lengthy text sections must be read the advantages of paper in terms of image quality lead to speed deficits with electronic text.

8.10 The Validity and Utility of the Framework

The framework has been used in two studies, involving two different texts and four different electronic applications. Its validity has been tested by parsing verbal protocols of readers into convenient chunks and relating them to the elements in the framework where they have been shown to map adequately and sufficiently. For the purposes of providing a non-complex representation of reader psychology relevant to text (rather than a formal psychological model) use it can be deemed valid.

The utility of the framework was tested by examining the accuracy of predictions derived from the framework. Over both studies, five hypotheses were tested, four of which resulted in the rejection of the null hypothesis. No more need be said of these except that they were strong, i.e., unidirectional, hypotheses. For the hypothesis that was not supported it is possible that experimental design factors are sufficient to explain the data. There appears to be no need to alter the framework to explain the findings. Thus as a means of predicting the likely performance of readers with electronic texts, the framework has demonstrated utility.

In the final chapter therefore, attention turns to the envisaged applications of this framework and the lessons that have been learned in this work for the development of electronic text systems.

CHAPTER 9

DESIGNING ELECTRONIC TEXT: CONCLUSIONS AND PROSPECTS

9.1 Introduction

The thesis set itself the primary aim of examining and subsequently describing the reading process in a manner that would support sensible analysis of the potential role of technology in this process. The framework outlined in chapter seven is intended to serve this purpose and in so doing provide a means of conveying appropriate human factors knowledge to designers of electronic text systems, the second aim of the thesis. The present chapter will review the work in the light of these aims and suggest areas for future research.

9.2 The framework as a description of the reading process

9.2.1 Describing reading at an appropriate level of abstraction

The process of reading has been subjected to continued examination by scientists from a variety of disciplines for over a century now. For all that effort, the process has still to be adequately described by any one discipline. Psychology has led the way in trying to understand the cognitive activities involved, while information science has concentrated on the more pragmatic issues of providing people with access to stored material. Educationalists, typographers and sociologists have all applied their discipline's tools and theoretical perspectives and while collectively, progress can be said to have been made, few researchers of reading would claim to have all the answers.

In the present context, the impact of advanced information technology on the reading process was identified as an issue worthy of investigation. With the impetus provided by electronic text in general, and hypertext in particular, this issue is becoming the focus of much attention and speculation. Current research on reading electronic text was reviewed and found to be both piecemeal and of little direct use to those responsible for designing these tools, primarily as a result of the narrow uni-disciplinary definitions of reading adopted by researchers and the resultant failure of any descriptive framework to provide a means to conceptualise the range of issues involved. The work in this thesis is an attempt to fill this descriptive vacuum.

It became obvious from examining the question of usability with respect to electronic texts that the variance in texts and tasks was likely to be of crucial importance in describing the reading process. In order to operationalise these factors in a reader-relevant form it was decided to examine people's perceptions of texts and their characteristic manner of using them. The first stage of this, the repertory grid analysis suggested that all texts are ultimately describable by readers in terms of three criteria: why they read them, how they read them, and what general type of information they contain.

The importance of these criteria lies in their ability to distinguish between texts according to usage factors and thereby classify material in a form that directly supports examination of the potential role of information technology in their use. This sets the classification criteria apart from any other text typology, the majority of which have attempted to classify material in terms of linguistic structures (e.g., de Beaugrande, 1980) from which mappings to electronic text design are difficult to make. In comparison to the few typologies that have been proposed in the context of discussions on electronic text, the classification proposed in chapter four is probably the only one to emerge directly from the examination of readers' own classifications.

When a text has been conceptualised in these terms one has a basic orientation from which to proceed in further describing the reading process appropriately. Thus, given any text, the three criteria can be used to elicit detailed information from readers on the type of tasks it is used for, their manner of interacting with it and the context of typical use. This can be done directly by a researcher or designer based on common-sense reasoning or, as shown in chapter five, more objectively through structured interviews and simulated performance with readers.

In this way, the reading process is initially conceptualised in terms of the text and task involved, hence the initial element of the descriptive framework: the task processor[1]. This immediately distinguishes it radically from cognitive psychological analyses of reading which in many ways can be seen as text and task independent. It also distinguishes the description from the type of conceptualisation offered in information science which is concerned with the range of texts but offers little insight into how individual readers actually use them once they have been located.

---

[1] Perhaps a more accurate term would have been the text processor or text *and* task processor. However the former is too likely to be confused with some computing or cognitive element in traditional reading models and the latter is cumbersome. The term task in this element's title is intended to convey the text-dependant nature of reading through the use of the Why, What and How criteria for task analysis.

The reliance on task analysis as a function of text classification promotes a level of description that can be seen as predominantly psychological in its concepts yet is atypical (in its breadth) of traditional psychological descriptions of reading. In relative terms it is a higher level of description than that provided by cognitive psychology but a lower level of description than that typically provided by information science. This is obvious from two other components of the descriptive framework: the information model and the manipulation facilities and skills element.

The concept of an information model is well-established in the psychological and linguistic literature but tends to be used only as a theoretical construct in discussions on reading comprehension (see e.g., van Dijk, 1980 or Garnham, 1987). The link between this work and the more traditional research on reading is only infrequently made. However, the interviews with readers carried out in this thesis confirm that the concept is inextricably interwoven with text usage, providing a reader with the means of grasping the organisation of material as well as supporting accurate prediction of the location of material in a text. The experimental work in chapter six on academic journal articles, and extracts from the verbal protocols of readers interacting with a unique text in chapter eight lend support to these views.

The manipulation element is perhaps the least likely component of the descriptive framework. Few people, when discussing reading, ever consider the issue of document manipulation to be of central (if any) importance. However, from the literature on reading from screen reviewed in chapter two, it was obvious that manipulation issues are crucial to the analysis of electronic text. Much reading involves manipulation by virtue of the presentation media humans have developed. From pamphlets to ledgers, letters to novels and manuals to encyclopedia, reading invariably requires the reader to open and turn pages, keep fingers in the text portion of interest while opening other sections and so forth. In fact, it is such an inextricable part of the process that without the ability to manipulate material easily, much reading would not be possible (or at best, would prove difficult) with current print media. The framework recognises the importance of these activities by including a manipulation element in its structure.

The lowest level of the framework represents behaviour more usually equated with the activity of reading. The serial reading element is the component that covers the process of extracting the message from the text, i.e., it refers to the contact or interaction between eye and print, so to speak. When an individual actually examines the text at the word or sentence level, the type of activities common to traditional psychological models of reading such as eye movement, word recognition, lexical processing and so

forth are presumed to occur. From the point of view of the framework, these issues are pertinent, but only after, or in conjunction with, the range of behaviours and processes described in the other elements.

The framework therefore describes reading as a task driven activity involving the setting of goals, the evolution and application of an information model, the manipulation of a document and the visual processing of text images. This is in contrast to the definition of reading as the visual and cognitive processing of textual images typical of psychological textbooks or as the acquisition and usage of documentation, to put it in information science terms. It does not suggest that these are the only issues that can be be validly described as reading, nor does it imply that any one of these is more/less important in the whole process. Furthermore, it does not suggest that traditional research paradigms on reading are wrong. Its intention is purely to provide a level of discourse appropriate to the examination of reading in the context of information technology.

9.2.2 The scope of the framework

Each of the elements in the framework raises an issue or set of issues to be dealt with in the design of electronic text. Thus the reading task must first be understood in the terms of the text type and its context of use. The information model element focuses attention on the reader's representation of the document's structure. The manipulation element highlights the importance of such facilities while the serial reading element raises the issues associated with visual ergonomics. Issues that do not map onto one or other of these elements are, according to this framework, of secondary importance to the design of electronic texts.

This latter point is worth elaborating. No scope for the *explicit* analysis of the reading outcome is provided by this framework. So, for example, the concept of comprehension, amongst others, is not represented in the framework; yet comprehension is, for many theorists, a crucial component of reading. This is not a return to the theoretical debate on the appropriateness or otherwise of comprehension in the discussion of reading with a statement of the present author's recommendation to exclude it. Rather it is a reflection of the goal of the descriptive framework: to support the accurate examination of human factors issues in electronic text design.

If technology is designed appropriately, users will be able to gain access to well presented information in an efficient and easy manner. At this point, it is not clear what

more can be done to ensure the reader actually makes sensible usage of this material, i.e., achieves his goal, finds his reference, comprehends the text and so forth. This choice of outcome exclusion emphasises the lack of applicable knowledge available from work on comprehension and similar concepts even for designers of paper texts, which implies that attempting to design electronic text that ensures greater comprehension of material for example is not an immediately measurable goal. Obviously, as cognitive science progresses, such goals might become more feasible. They are certainly desirable. However, the present author's view is that currently, they are not practicable, in the sense that a design process cannot be specified sufficiently to ensure their attainment. [2]

Furthermore, it is highly unlikely that one level of description can hope to encompass all possible issues. The nature of scientific investigation is that it divides the world up into levels of analysis. Unified theories are rare (even within religions) and classical science divides itself into the disciplines of physics, chemistry and biology, none of which attempts explanations outside its accepted boundaries. It should not be expected therefore that given current knowledge, one could devise a single level description of such a complex human activity as reading.

Until we have sufficient knowledge about the relationship between information presentation and subsequent learning or comprehension, then the efforts of electronic text designers should be concentrated on providing the tools to access and manipulate relevant material in a suitable manner. This is not defeatist or pessimistic however. The attainment of comprehension or other outcomes are likely to be contingent upon such successful and easy access provided by well-designed systems i.e., such well-designed systems are likely to result in greater (or at least faster) comprehension than badly designed ones. In this sense the consideration of such issues is not dismissed but is placed in perspective. The reading process as described in this framework is surely a prerequisite to any desirable outcomes such as comprehension. The present framework's exclusion of such concepts from its description of immediately relevant issues is not a dismissal of them but a recognition of their complexity.[3]

---

[2] Interestingly, not all cognitive scientists consider comprehension to be an issue worth addressing. For example, van Dijk and Kintsch (1983) stated:

> "there is no unitary process 'comprehension' that could be measured once and for all if we could but find the right test. Comprehension is a common sense term which dissolves upon closer analysis into many different sub-processes. Thus we need to construct separate measurement instruments for macroprocesses, knowledge integration, coherence, parsing .....Comprehension is just a convenient term for the aggregation of these processes: it is not to be reified, not to be tested for" (p.260).

[3] Put simply, one would not expect a specification of a product to state that it must be built so as to ensure greater comprehension. Though this might be desirable or even required, the specification would

9.2.3 The framework as context provider for research findings

The framework can also be seen as an aid to understanding the human factors literature on reading. As described in chapter two, this literature is replete with empirical studies on issues such as the effect of image polarity, scrolling versus paging, large versus small screens and so forth. Interpretation of the various findings can prove problematic and there are contradictions in findings that can not be resolved without reference to contextual factors.

The framework offers such a context within which to assess the findings of any one experiment. Thus when one is presented with the question of optimum screen size and notes the Elkerton and Williges (1984) finding that there is no significant difference between screen sizes of 5 lines and anything larger and contrasts this with Dillon *et al.*'s (1990a) conclusion that screens of 60 lines result in significantly different manipulations than, and are preferred over, 20 line screens, the framework supports the interpretation of each of these findings in an appropriate context by suggesting how variables relating to tasks, texts, manipulation etc. must be considered. Likewise, when Gould *et al.* (1987b) claim that reading from screen can be as fast as reading from paper, the framework supports an interpretation of this statement that allows an informed (i.e., non-literal) acceptance.[4] The simple heuristic being: for any statement about the advantages or disadvantages of electronic text, consider its reflection of each of the four elements in the framework. If it misses one (i.e, fails to include assessment of each element) then it is incomplete.

The issue of statement completion is interesting and useful. To make a complete statement about electronic text reference must be made to the task, the text, the reader's model, the type of manipulation facilities available and the visual ergonomics. For example, the statement:

> paper is better for proofreading tasks than electronic text;

is less complete than the statement: ,

---

state it in more concrete terms e.g., the system must be faster, more accurate etc., variables which are presumed to reflect, lead to, or correlate with comprehension.

[4] In this case the framework highlights the fact that electronic text can be as fast for proofreading short texts on an ergonomically optimised screen. However this does not mean that no speed deficits occur for other tasks or texts even with such optimised screens.

> for proofreading a familiar text form, paper is better than electronic text.

Though both of these are less complete than the statement:

> for proofreading a familiar text form, on a typical screen with scrolling facilities, paper is better than electronic text.

In each of these cases the references to particular components of the framework are easily seen. However, despite a statement's completion, its truth content is another factor. A statement may be complete in the sense implied here, but be wrong. However, this is a separate issue. A complete statement is open to evaluation, either in terms of current knowledge or empirical investigation. An incomplete one cannot be so easily tested. For example, the third statement above, is easier to comment appropriately on or empirically demonstrate as valid than the first statement. It befalls researchers and designers alike therefore, when making claims about electronic text, to do so in as complete a fashion as possible. Similarly, incomplete statements should be treated with caution.

9.3 The framework as a guide to designers

9.3.1 Why a qualitative framework?

At the outset it was stated that a secondary aim of the thesis was to ensure that any resulting description could be packaged and presented in a form suitable for use at the earliest stages of design. It is intended that the framework as outlined in chapter seven, also satisfies this criterion.

The delivery format of a qualitative framework rather than any alternative such as a set of guidelines or a quantitative model was adopted for a variety of reasons. Predominantly, the emerging perspective on reading was not amenable to reliable quantification. There are few aspects of reading and information design that are amenable to such analysis and the reception of other quantitative models of HCI (e.g., the GOMS model of text editing) by the design community at large hardly inspires confidence in their applicability.[5]

---

[5] The obvious exception here would be the visual ergonomics issues for which standards on resolution, luminance etc. can be stated quantitatively. Particular aspects of manipulation might also be quantifiable (see. e.g., Card et al 1978). However, these are very specific instances of HCI that are not singularly concerned with reading.

Guidelines were not adopted as there are several problems with them that are well documented, not least their inherent contradictions and over-generalisations. As Hammond *et al.* (1987) put it :

> "If behaviour results from an interplay of factors, so will the ease of use of an interface. These interdependencies are hard, or even impossible, to capture in simple statements. A guideline which is true in one context may well be misleading in another....the more complex the interface, the less plausible it is that guidelines will help" (p. 41).

The qualitative framework is seen as a suitable alternative to both the standard models or guidelines approaches. It represents a stylistically simple way of presenting a set of complex ideas and supports "unpacking" of the components to facilitate more detailed analysis. By representing reading as the interaction of a small number of elements it focuses attention on the range of issues to be considered and their possible inter-relationships.

The term "unpacking" is meant to imply that other forms of advice could be derived from a framework such as this. Guidelines for example could be "unpacked" from particular components, e.g., "when transferring paper to hypertext, retain the useful structural components of the original" could be a guideline derived from the IM element, or "for presenting text on screen ensure image quality is high" could be similarly derived from the SRP element.

Alternatively, existing guidelines could be interpreted in the light of the framework to ensure contextual issues are addressed (thereby lessening one of the major shortcomings of guidelines– their over generalised form). For example, the guideline: "when displaying text that will not fit on a single screen, then use paging rather than scrolling" (Rivlin *et al.*, 1990), if applied rigidly, would lead to some very unusable designs. But if interpreted in the context suggested by the framework i.e., for certain users, doing particular tasks with specific texts, it is unlikely to be followed slavishly (and, ultimately, inappropriately) by a well meaning designer. [6]

---

[6] The Rivlin et al guidelines are a prime example of the problems inherent in such advisory formats. While they provide generally useful information to designers, the published set contains at least two erroneous suggestions and several, like the cited example, which sound authoritative but generally fail to allow for important contextual variables which negate their recommendation. For a further review of these guidelines see Dillon (1990).

9.3.2 Potential uses of the framework by designers

As well as being the most suitable presentation format, the framework is intended to support several uses. First, a designer could use it simply as a checklist to ensure that all important components of the text under design are considered. This guards against the reliance on research findings at one level to ensure good design (e.g., just following the advice on visual ergonomics which concludes that certain fonts, polarity and resolution variables can overcome the reading speed deficit). While that advice might be pertinent and applicable, the framework would suggest that it is but one part of the design problem.

Second, it could be used to guide design by allowing a designer to conceptualise the issues to be dealt with in advance of any specification or prototype. In this sense its advocated use is as an advance organiser (Ausubel, 1968), enabling the designer to organise his thoughts on the problem and highlight attributes of the specification that need to be considered. As discussed in chapter seven, such an application could lead to significantly more appropriate first specifications or prototypes, lessening the number of iterations required and thereby reducing the time and costs involved in design.

Third, the framework supports the derivation of predictions about readers' performance with a document. The uses made of the framework in the previous chapter highlight its value as a predictive tool for a human factors practitioner, adequately familiar with the research in this area, to predict the type of problems a reader will face using an electronic document. It is the author's view that all of the predictions made were easily derived from the framework through the analysis of the various elements and their manifestation or support in the relevant designs, and that few practitioners would face difficulties deriving similarly accurate predictions in other text/task environments.

Finally, the framework has potential evaluative applications. It could be used to guide expert evaluation of a system under development (i.e., a usability assessment) and support troubleshooting for weaknesses in design. This proposed use is not unlike the first use outlined above except it occurs at a different stage in the design process and is intended to support reasoned examination of the quality of an instantiated design. In this role, one could imagine a designer using the framework to check how the system rated on variables such as image quality, the information model it presents, the type of tasks it will support or manipulations it enables.

9.3.3 Actual use of the framework at HUSAT

The framework was derived as a result of the successful input of the results of several studies carried out in this thesis to the hypertext journal database developed at HUSAT (see McKnight *et al.*, 1990b). The repertory grid analysis of text classifications provided the appropriate criteria for examining academic journals. This examination was the journal usage study reported in chapter five. That study highlighted the importance of the reader's perception of structure in documents and raised the issue of article structure in the hypertext database under development. This issue was examined empirically in the studies reported in chapter six. The proposed framework is therefore a representation of the issues found to be of importance to the design team building a hypertext system.

The framework has subsequently formed the basis of a usability requirements specification carried out at HUSAT for a system called CHIRO.[7] The present author used the framework to structure initial investigations into user requirements and task performance with the paper database. A Why, What and How analysis of document use was carried out and a usability specification for the system derived. Application of the framework also highlighted the need to examine more closely the information model the users would possess. Details of this work are reported in McKnight *et al.* (1990c). The framework is also likely to be used in any subsequent evaluations of the CHIRO database. For present purposes, it highlights the type of usage of the framework that it is hoped other designers will be able to make. This is an obvious area for future work.

9.3.4 ADONIS: a reprise

In chapter three, the ADONIS document supply system was described and used to highlight many of the shortcomings the author felt to be present in the literature professing to advise designers i.e., the system reflected piecemeal application of some guidelines while totally missing some important aspects of the reading scenario that tend not to be covered in the literature. It is reasonable to ask therefore if the proposed framework would have been of any more use to the ADONIS designers.

It goes without saying that this question cannot be answered entirely satisfactorily,

---

[7] CHIRO stands for Collaborative Hypertext In Research Organisations and is a British Library-funded project investigating the design and use of a multi-user hypertext database. This database will replace a large collection of paper documents used by a research team.

without giving the framework to the designers and letting them specify an alternative system while controlling for any further knowledge they might have acquired in the interim. However, one can openly speculate on how ADONIS might have been different if the design team had considered the framework.

Given that the major problems with ADONIS lay with its poor manipulation facilities, it is tempting to assume that even a cursory examination of the issues presented in the framework would have led to more consideration of these in the design team. Considering that the someone involved in the design was obviously concerned enough about image quality to provide a high resolution, black on white screen (for the most part) the framework would have made it clear that this was only one level of issue out of four and therefore, some attention should be paid to the other levels. Had this been done, it is unlikely that a design which restricted manipulation so much, or made searching so awkward would have emerged.

Obviously this is speculation and is not worth pursuing further, but it provides an image of how this framework might be employed. The earliest stages of creative thought are not well understood. In the time it takes to produce an initial idea or respond to a request for a specification the designer must apply a body of knowledge that is a mixture of intuition, experience, stored facts and opinions (Dillon and Sweeney, 1988). If the framework can lead to appropriate inputs to even one of these knowledge sources then it is likely to have done some benefit, and a reader utilising an electronic text at some future time will reap the rewards of a usable system.

9.4. Further research

Several aspects of the framework have been highlighted as worthy of further investigation. In particular not enough is known about the characteristic manner of reading involved for particular texts or text/task combinations. The criteria outlined in the repertory grid study and subsequently applied to the analysis of journals and manuals could be usefully employed to this end. Such a classification of a wide variety of texts would be useful and interesting for those concerned with electronic text design.

A specific topic of relevance is the existence of information models for the type of texts likely to find their way into the electronic medium. The work reported in this thesis has concentrated mostly on journal articles (and a specific subset thereof), yet it is argued that the concept of structural models is relevant for a multitude of texts. This would

then be a natural area for further work and one that cannot be avoided if hypertext is to become anything more than a research curiosity which, one could argue, is all it is at this time.

In terms of basic human factors work there is still a lot to be learned about screen ergonomics for electronic text, particularly in some of the areas outlined in the literature review such as screen size, manipulation facilities, search facilities and icon design. These are all current research themes in the human factors discipline but some specific work aimed at the electronic text domain would be useful as it is not always clear how findings from one area of application transfer to another.

In contrast to these areas which are on-going concerns of the hypertext and human factors research communities, the descriptive framework proposed in this thesis needs to be tested with other designers building a variety of text systems. This would serve to identify not only its particular value but the utility of qualitative frameworks in general for hypertext design. Unlike the other investigations of various reader-text issues outlined above, this is a task that could not be dealt with by simple experimental means or short studies. The type of knowledge that is required would ideally require the investment of large amounts of time and resources, of the type more normally given over to government funded projects under intiatives such as ESPRIT or RACE.[8] In the absence of such resources the framework could be tested through a number of small scale projects with different designers. This is certainly an intention of the author, but it rests ultimately on the goodwill of others rather than the enthusiasm of the author.

9.5 Specifying the design process for hypertexts

At the end of a thesis such as this it is justifiable to ask if the work could be summarised into some applicable advice for designers. While it is the intention of the author that the framework should fulfil this requirement, the thesis is capable of supporting a more explicit statement of how electronic text should be designed to ensure usability. This final section provides a design sequence that involves the framework's components and offers, on the basis of the author's experience, a good chance of successful goal attainment.

---

[8] The ESPRIT funded project HUFIT (Human Factors in Information Technology) was one such project (see e.g.Galer and Taylor 1989). It ran for five years, involved several major IT suppliers, cost more than £1millon and was intended to produce human factors tools for designers. Interestingly, the actual goal of the project was to design and deliver such tools. The uptake of them amongst IT companies outside of the consortium was considered beyond the scope of such a project, which gives an indication of the amount of work required to input such tools into real-world design teams.

Designing a usable hypertext database therefore involves the following stages:

- Task analysis of the text involved according to three criteria: How it is used, Why it is used and What readers perceive the information to be;

- Investigate the extent to which the document structure is fixed by existing readers' models;

- Determine the electronic structure by considering the readers models and the tasks being performed;

- Consider the manipulation facilities required for basic use and ensure that readers can at least perform these activities with the mechanisms provided;

- Attempt to add value to the system by offering facilities to perform activities impossible or difficult on paper;

- Ensure image quality is high;

- Test system on users performing real tasks;

- Re-design accordingly.

The first step is important and will provide information of direct relevance to the next three steps. The last step is probably the most important although it is often seen as a luxury that cannot be afforded. Failure to test the design is bound to lead to problems as no theoretical models or formal guidelines exist that can even approximate the quality of information obtained from observing real users interacting with a system. This applies to the descriptive framework proposed in this thesis as much as to any existing model in HCI. These steps will not guarantee success but they offer better prospects of achieving it than any others.

9.6 General Conclusion

In 1908, Edmund Huey wrote that to understand reading would be the acme of the

psychologist's achievements. That statement is perhaps seen to be more accurate with each successive generation of research on the subject. The subtlety and complexity of the reading process makes it a taxing problem for anyone intent on examining it.

The present work has carved out but part of the reading process as its subject matter. In so doing it has drawn on the ideas and themes of several disciplines concerned with reading. While it might have appeared critical, particularly of the work in cognitive psychology, it is within the psychological perspective that the present work most appropriately lies. It cannot claim to have explained the process satisfactorily, or to have solved any of the thorny issues of what humans do when they read texts. However it has led to a perspective; one that aims at improving the quality of the reading process and ensuring that technology does not make us read despite itself, but actively supports us in this quintessentially human activity. The question is not 'should documents be paper or electronic?' but 'how can any medium best satisfy an information need?' There is no simple answer but the framework can help us better understand the question.

9.7. The prospects for electronic text

In many ways it seems as if the role of human factors studies of electronic text has been to highlight problems with the technology, to show that paper is inherently better or to criticise designers and advocates of the medium for overlooking the human issues. It is hoped that this thesis has not presented a singularly negative view of electronic text but a realistic one, tempered with the optimism that comes from the author's belief that good design is both possible and beneficial. This section, the final one of the thesis, examines the prospects for electronic text in the light of the work reported.

Almost ten years have passed since Jonassen (1982) uttered the memorable (and now punishable by ridicule) phrase:

> "in a decade or so, the book as we know it will be as obsolete as is movable type today" (p. 379).

Whatever the facts about movable type in 1982, the book as we know it is certainly far from obsolete in the early 1990s. Jonassen is not alone; the advent of hypertext and desk-top computing means that his point of view is considered visionary in some quarters and that the truth of his claim lies not in its timescale but its implications.

The implications of widespread electronic text "any year now" are important. As this thesis has attempted to highlight, documentation is everywhere: at home in the form of anything from instructions for operating microwave ovens to the novels that induce sleep; at work in the form of texts ranging from reports on latest developments in company sales to the memos that descend from above; and in the world at large in the form of newspapers, advertising boards, shop catalogues and so on. Avoiding documentation in contemporary industrial societies would be a feat of Herculean proportions. Modifying documentation therefore, by presentation in electronic rather than paper forms, will undeniably have an impact on our lives.

In this light the zeal of advocates is understandable, it's just that when humans enter into the equation, accurately predicting these impacts becomes difficult. Paper is familiar, is well liked, easy to use (most of the time), affords a representation of its structure that is quickly acquired by readers and leads to the emergence of conventional forms, is portable, supports excellent image quality and is cheap since publishers have long since recovered their capital outlay on production equipment. Obviously examples could be found of paper documents which flout such conventional benefits but they hold true for the majority of paper texts, while the reverse tends to hold for electronic ones.

According to the findings of this thesis, the progress of electronic text will be neither explosive nor all-embracing. It will only progress by gaining footholds in small task and text domains and by being found usable there (and possibly, at first, only by a few enthusiasts in these domains). As technology develops, screens improve, portability increases and resistance is lowered the scope for electronic text will broaden, but there is little reason to believe paper will become obsolete in the near future (if ever).

The process will be accelerated by good design, of the kind advocated in this thesis, but conversely, it will be hampered by weak design i.e., that which fails to consider all elements of the framework. It is highly unlikely that there is anything inherently constraining in the concept of electronic text that cannot be solved by technological improvements and increased knowledge of human reading. However, the process of reading is not simple and texts are used in multiple ways for myriad tasks by millions of people. Perhaps the only reasonable prediction that can be made is that we shall witness the emergence of dual-form documents: electronic versions for some tasks, paper versions for others. The strengths of the computer will enable cheap storage and rapid access while the intimacy and familiarity of paper will be retained for detailed studying and examination of material.

A text without a reader is worthless. Similarly, a technology without a user is pointless. The human is the key; only by relating technologies to the needs and capabilities of the user can worthwhile systems be developed. The work in this thesis is a step in that direction for electronic texts, but there is a long journey ahead.

REFERENCES

Anastasi, A. (1990) *Psychological Testing* (Sixth Edition). New York: Macmillan.

Anderson, J. (1980) *Cognitive Psychology and Its Implications*. San Francisco: W.H. Freeman.

Anderson, J. (1983) *The Architecture of Cognition*. Cambridge MA: Harvard University Press.

Askwall, S. (1985) Computer supported reading vs reading text on paper: a comparison of two reading situations. *International Journal of Man-Machine Studies*, 22, 425-439.

Ausubel, D.P. (1968) *Educational Psychology: a Cognitive View*. New York: Holt, Rinehart and Winston.

Bannister, D. and Fransella, F. (1971) *Inquiring Man*. Harmondsworth: Penguin.

Bannister, D. and Mair, D. (1968) *The Evaluation of Personal Constructs*. London: Academic Press.

Barber, P. (1988) In favour of theory. *Ergonomics*, 31(6) 871-872.

Bartlett, F. C. (1932) *Remembering*. Cambridge: Cambridge University Press.

Bauer, D. and Cavonius, C. R. (1980) Improving the legibility of visual display units through contrast reversal. In E. Grandjean and E. Vigliani (eds.) *Ergonomic Aspects of Visual Display Terminals*. London: Taylor and Francis.

Bauer, D., Bonacker, M. and Cavonius, C.R. (1983) Frame repetition rate for flicker-free viewing of bright VDU screens. *Displays*, January, 31-33.

de Beaugrande, R. (1980) *Text, discourse and process*, Norwood NJ: Ablex.

de Beaugrande, R. (1981) Design criteria for process models of reading. *Reading Research Quarterly*, 16(2), 261-315.

Beeman, W., Anderson, K., Bader, G., Larkin, J., McClard, A., McQuillan, M. and Shields, M. (1987) Hypertext and pluralism: from lineal to non-lineal thinking. In: *Proceedings of Hypertext '87.* University of North Carolina, Chapel Hill, 67-88.

Beldie, I.P., Pastoor, S. & Schwartz, E. (1983) Fixed versus variable letter width for televised text. *Human Factors,* 25(3), 273-277.

Belmore, S. (1985) Reading computer presented text. *Bulletin of the Psychonomic Society,* 23(1), 12-14.

Benest, I.D. (1989) A hypertext system with controlled hype. In. R. McAleese and C. Green (eds.) *Hypertext: State of the Art.* Oxford: Intellect. 52-63.

Bertaggia, N., Novarra, F., Dillon, A. and Bonner, J. (1987) Application of the methodology to the PLC and its impact. Working Paper A5.3. HUFIT Ref 4-12/87.HUSAT Research Institute, Loughborough.

Bevan, N. (1981) Is there an optimum speed for presenting text on VDUs. *International Journal of Man-Machine Studies,* 14, 59-76.

Billingsley, P. (1982) Navigation through hierarchical menu structures: does it help to have a map? *Proceedings of the Human Factors Society 26th Annual Meeting.* Santa Monica CA: Human Factors Society, 103-107.

Binder, A. (1964) Statistical theory. In: P. Farnsworth, O. McNemar and Q. McNemar (eds.) *Annual Review of Psychology 1964,* 15, 277-310.

Brems, D. and Whitten W. (1987) Learning and preference for icon-based interfaces. In: *Proceedings of the 31st Annual Meeting of the Human Factors Society.* Santa Monica CA: Human Factors Society, 125-129.

Brewer, W. (1987) Schemas versus mental models in human memory.In: I.P. Morris (ed.) *Modelling Cognition.* London: John Wiley and Sons, 187-197.

Brown, P. (1988) Hypertext: the way forward. In: J.C. Van Vliet (ed.) *Document Manipulation and Typography.* Cambridge: Cambridge University Press, 183-191.

Buckley, P. (1989) Expressing research findings to have a practical influence on design. In J. Long and A. Whitefield, (eds.) *Cognitve Ergonomics and Human Computer Interaction.* Cambridge: Cambridge University Press, 166-190.

Buie D. and Hassell, J. (1982) An Introduction to Wines. HyperTIES Demonstration Document. Distributed with HyperTIES Application Program (v.1986). Ben Shneiderman, University of Maryland.

Bush, V. (1945) As we may think. *Atlantic Monthly*, 176(1), 101-108.

Cakir, A., Hart, D. J. & Stewart, T. F. M. (1980) *Visual Display Terminals.* Chichester: John Wiley and Sons.

Canter, D. (1984) Wayfinding and signposting: penance or prosthesis? In: R. Easterby and H. Zuraga (eds.) *Information Design.* Chichester: John Wiley and Sons, 245-264.

Canter, D., Rivers, R. and Storrs, G. (1985) Characterizing user navigation through complex data structures. *Behaviour and Information Technology* 4(2), 93-102.

Card, S., English, W. and Burr, B. (1978) Evaluation of mouse, rate-controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics*, 21, 601-613.

Card, S.K., Moran, T.P. and Newell, A., (1983) *The Psychology of Human-Computer Interaction.* Hillsdale NJ: Lawrence Erlbaum Associates.

Carroll, J. (1984) Minimalist design for active users. In: B. Shackel (ed.) *INTERACT '84.* Amsterdam: North-Holland, 39-44.

Carroll, J. (1990) Infinite detail and emulation in an ontologically minimized HCI. In: *CHI'90.* NewYork: Association of Computing Machinery, 321-327.

Carroll, J. and Campbell, R. (1986) Softening up hard science: a reply to Newell and Card. *Human-Computer Interaction*, 2(3), 227-249.

Carroll, J. and Thomas, J. (1982) Metaphor and cognitive representation of computing systems. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-12(2), 107-116.

Catterall, B., Allison, G. and Maguire, M. (1989) HUFIT: Specification and Design Tools. In E.D. Megaw (ed.) *Contemporary Ergonomics '89*. London:Taylor and Francis, 97-102.

Chalmers, A. (1976) *What is this thing called science?* Milton Keynes: Open University Press.

Chapanis, A. (1988) Some generalizations about generalizations. *Human Factors*, 30(3) 253-268.

Chapman, L. J. and Hoffman, M. (1977) *Developing Fluent Reading*. Milton Keynes: Open University Press.

Charney, D. (1987) Comprehending non-linear Text: the role of discourse cues. In: *Proceedings of Hypertext '87*. University of North Carolina, Chapel Hill, 109-120.

Cohen, G. (1988) *Memory in the Real World*. London: Lawrence Erlbaum Associates.

Coltheart, V. and Evans, J. (1982) An investigation of semantic memory in individuals. *Memory and Cognition*, 9(5), 524-532.

Conklin, J. (1987) Hypertext: an introduction and survey. *Computer*, September, 17-41.

Creed, A., Dennis, I. & Newstead, S. (1987) Proof-reading on VDUs. *Behaviour and Information Technology*, 6(1), 3-13.

Crowder, R. (1982) *The Psychology of Reading: An Introduction*. New York: Oxford University Press.

Cushman, W. H. (1986) Reading from microfiche, VDT and the printed page: subjective fatigue and performance. *Human Factors*, 28(1), 63-73.

van Dijk, T.A. and Kintsch, W. (1983) *Strategies of Discourse Comprehension.* London: Academic Press.

van Dijk, T.A. (1980) *Macrostructures.* Hillsdale NJ: Lawrence Erlbaum Associates.

Dillon, A. (1987) Knowledge acquisition and conceptual models: a cognitive analysis of the interface. In: D. Diaper and R. Winder (eds.) *People and Computers III.* Cambridge: Cambridge University Press, 371-379.

Dillon, A. (1988) The ADONIS Document Delivery Workstation: a user interface evaluation. Project Quartet Deliverable, HUSAT Research Centre, Loughborough University of Technology.

Dillon, A. (1990) A review of Rivlin et al (1990): Guidelines for Screen Design. *Hypermedia,* 2(2), 171-173.

Dillon, A. and Sweeney, M. (1988) The application of cognitive psychology to CAD. In D.M. Jones and R. Winder (eds.) *People and Computers IV.* Cambridge: Cambridge University Press, 477-488.

Dillon, A., Richardson, J. and McKnight, C. (1990a) The effect of display size and paragraph splitting on reading lengthy text from screen. *Behaviour and Information Technology,* 9(3), 215-227.

Dillon, A., Richardson, J. and McKnight, C. (1990b) Navigation in hypertext: a critical review of the concept. In D. Diaper, D. Gilmore, G. Cockton and B. Shackel (eds.) *INTERACT'90.* North Holland: Amsterdam, 587-592.

Dillon, A., Sweeney, M., Herring, V., John, P. and Fallon, E. (1988) The psychology of designer style. *The Alvey Conference 1988.* DTI/IED Publications, 271-275.

Downs, R. and Stea, D. (eds.) (1974) *Image and Environment: Cognitive Mapping and Spatial Behaviour.* London: Edward Arnold.

Duchnicky, R.L. and Kolers P.A. (1983) Readability of text scrolled on a visual display terminal as a function of window size. *Human Factors,* 25(6), 683-692.

Dunn, R. (1984) *Software Defect Removal.* New York: McGraw-Hill.

Eason, K. (1988) *Information Technology and Organisational Change.* London: Taylor and Francis.

Easteal, C. and Davies , G. (1989) *Software Engineering: Analysis and Design.* London: McGraw-Hill.

Edwards, D. and Hardman, L. (1989) "Lost in Hyperspace": cognitive mapping and navigation in a hypertext environment. In: R. McAleese (ed.) *Hypertext: Theory into Practice.* Oxford: Intellect, 105-125.

Egan, D., Remde, J., Landauer, T., Lochbaum, C. and Gomez, L. (1989) Behavioural evaluation and analysis of a hypertext browser. *Proceedings of CHI'89.* New York: Association of Computing Machinery, 205-210.

Elkerton, J. and Williges, R. (1984) Information retrieval strategies in a file search environment. *Human Factors,* 26(2), 171-184.

Elm, W. and Woods, D. (1985) Getting lost: a case study in interface design. *Proceedings of the Human Factors Society 29th Annual Meeting.* Santa Monica CA: Human Factors Society, 927-931.

Engelbart, D. (1963) A conceptual framework for the augmentation of man's intellect. In: P. Howerton and D. Weeks (eds.) *Vistas in Information Handling,* Vol. 1. London: Spartan Books, 1-29.

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

Ewing, J., Mehrabanzad, S., Sheck, S., Ostroff, D. and Shneiderman, B. (1986) An experimental comparison of a mouse and arrow-jump keys for an interactive encyclopedia. *International Journal of Man-Machine Studies,* 24(1), 29-45.

Eysenck, M. (1983) *A Handbook of Cognitive Psychology.* London: Lawrence Erlbaum Associates.

Feldman, T. (1990) The emergence of the electronic book. British National Bibliography Research Fund Report 46, The British Library.

Ferguson, G. (1959) *Statistical Analysis in Psychology and Education.* New York: McGraw-Hill.

Galer, M. and Taylor, B. (1989) Human Factors in Information Technology: Esprit Project 385. In: E. Megaw (ed.) *Contemporary Ergonomics 1989.* London: Taylor and Francis, 82-86.

Gardiner, M. and Christie, B. (eds.) (1987) *Applying Cognitive Psychology to User-Interface Design.* Chichester: John Wiley and Sons.

Garland, J. (1982) *Ken Garland and Associates: Designers-20 years work and play.* Cited in, Waller, R. (1987) The typographic contribution to language: towards a model of typographic genres and their underlying structures. PhD Thesis, Dept. of Typography and Graphic Communication, University of Reading.

Garnham, A. (1987) *Mental Models as Representations of Text and Discourse.* Chichester: Ellis Horwood.

Gittens, D. (1986) Icon based human-computer interaction. *International Journal of Man-Machine Studies,* 24, 519-543.

Gordon, S., Gustavel, J., Moore, J. and Hankey, J. (1988) The effects of hypertext on reader knowledge representation. *Proceedings of the Human Factors Society 32nd Annual Meeting* Santa Monica CA: Human Factors Society, 296-300.

Gould, J. D. & Grischkowsky, N. (1984) Doing the same work with hard copy and cathode-ray tube (CRT) computer terminals. *Human Factors,* 26(3), 323-337.

Gould, J.D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N. and Minuto, A. (1987a) Reading is slower from CRT displays than from paper: attempts to isolate a single variable explanation. *Human Factors,* 29(3) 269-299.

Gould, J.D., Alfaro, L., Finn, R., Haupt, B. and Minuto, A. (1987b) Reading from CRT displays can be as fast as reading from paper. *Human Factors,* 29(5), 497-517.

Hammond, N., Jorgensen, A., Maclean, A., Barnard, P. and Long, J. (1983) Design practice and interface usability: evidence from interviews with designers. IBM Hursley Human Factors Report HF 082, Hursley Park, Winchester.

Hammond, N. and Allinson, L. (1989) Extending hypertext for learning: an investigation of access and guidance tools. In: A. Sutcliffe and L. Macaulay (eds.) *People and Computers V*. Cambridge: Cambridge University Press, 293-304.

Hammond, N. and Allinson, L. (1987) The travel metaphor as design principle and training aid for navigating around complex systems. In: D. Diaper and R. Winder (eds.) *People and Computers III*. Cambridge: Cambridge University Press, 75-90.

Hammond, N., Gardiner, M., Christie, B. and Marshall, C. (1987) The role of cognitive psychology in user-interface design. In: M. Gardiner and B. Christie (eds.) *Applying Cognitive Psychology to User-Interface Design*. Chichester: John Wiley and Sons, 13-52.

Hannigan, S. and Herring, V., (1986) The role of human factors in the design of IT products. Deliverable A12b. ESPRIT Project 385 - HUFIT. HUSAT Research Institute, Loughborough.

Harrison, M. and Thimbleby, H. (eds.) (1990) *Formal Methods in Human Computer Interaction* . Cambridge: Cambridge University Press.

Hartley, J. (1985) *Designing Instructional Text* (2nd ed.). London: Kogan Page.

Hassard, J. (1988) FOCUS as a phenomenological technique for job analysis: its use in multiple paradigm research. *International Journal of Man-Machine Studies*, 27, 413-433.

Hatt, F. (1976) *The Reading Process* London: Clive Bingley.

Hayes-Roth, B. (1983) The blackboard architecture: a general framework for problem-solving? HPP Report No. HPP-83-30. Stanford University, Dept. of Computer Science.

Helander, M. G., Billingsley, P. A. & Schurick, J. M. (1984) An evaluation of human factors research on visual display terminals in the workplace. In: F. Muckler (ed.) *Human Factors Review: 1984.* Santa Monica CA: Human Factors Society, 55-129.

Huey, E. B. (1908) *The Psychology and Pedagogy of Reading.* New York: Macmillan.

Jaschinski-Kruza, W. (1990) On the preferred viewing distances to screen and document at VDU workplaces. *Ergonomics*, 33(8),1055-1063.

Johnson-Laird, P. (1983) *Mental Models* Cambridge: Cambridge University Press.

Jonassen, D. (1982) *The Technology of Text. Vol I. Principles for Structuring, Designing, and Displaying Text.* Englewood Cliffs NJ: Educational Technology Publications.

Jones, W.P. and Dumais, S.T. (1986) The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Transactions on Office Information Systems*, 4(1), 42-63.

Just, M.A. and Carpenter, P. (1980) A theory of reading: from eye movements to comprehension. *Psychological Review*, 87(4), 329-354.

Kak, A. V. (1981) Relationships between readability of printed and CRT-displayed text. *Proceedings of Human Factors Society - 25th Annual Meeting.* Santa Monica CA: Human Factors Society, 137-140.

Kelly, G. (1955) *The Psychology of Personal Constructs,* 2 Vols. New York: Norton.

Kerlinger, F. (1973) *Foundations of Behavioral Research.* New York: Holt, Rinehart and Winston.

Kerr S. T. (1986) Learning to use electronic text: an agenda for research on typography, graphics, and interpanel navigation. *Information Design Journal,* 4(3), 206-211.

Kieras, D. and Polson, P. (1985) An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies*, 22, 365-394.

Kintsch, W. (1974) *The Representation of Meaning in Memory* Hillsdale NJ: Lawrence Erlbaum Associates.

Kintsch, W. and Yarborough , J. (1982) The role of rhetorical structure in text comprehension. *Journal of Educational Psychology*, 74, 828-834.

Kline, P. (1988) *Psychology Exposed, or The Emperor's New Clothes* London: Routledge.

Kolers, P. A., Duchnicky, R.L. & Ferguson, D.C. (1981) Eye movement measurement of readability of CRT displays. *Human Factors*, 23(5), 517-527.

Kruk, R.S. & Muter, P. (1984) Reading continuous text on video screens. *Human Factors*, 26(3), 339-345.

Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Landauer, T. (1987) Relations between cognitive psychology and computer systems design. In: J. Carroll (ed.) *Interfacing Thought* : Cambridge MA: MIT Press, 1-25.

Lawson, B. (1979) *How Designers Think*. London: Architectural Press.

Licklider, J. (1965) *Libraries of the Future*. Cambridge MA: MIT Press.

Lovelace, E.A. and Southall, S.D. (1983) Memory for words in prose and their locations on the page. *Memory and Cognition*, 11(5), 429-434.

Mack, R.L., Lewis, C.H. and Carroll, J.M. (1983), Learning to use word-processors: problems and prospects. In: *ACM Transactions on Office Information Systems*, 1(3), 254-271.

Mandler, J. and Johnson, N. (1977) Remembrance of things parsed: story structure and recall. *Cognitive Psychology*, 9, 111-151.

Martin, A. (1972) A new keyboard layout, *Applied Ergonomics*, 3(1), 42-51

McAleese, R. (1989) Navigation and browsing in hypertext. In: R. McAleese (ed.) *Hypertext: Theory into Practice.* Oxford: Intellect, 6-44.

McClelland, J.L. and Rumelhart, D. (1981) An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review*, 88, 375-407.

McKnight, C. (1981) Subjectivity in sentencing. *Law and Human Behavior*, 5(2/3), 141–147.

McKnight, C., Dillon, A. and Richardson, J. (1990a) A comparison of linear and hypertext formats in information retrieval. In: R. McAleese and C. Green (eds.) *Hypertext: State of the Art.* Oxford: Intellect, 10-19.

McKnight, C., Dillon, A. and Richardson, J. (1990c) Project CHIRO: a progress report. HUSAT Research Institute, Loughborough University of Technology.

McKnight, C., Richardson, J. and Dillon, A. (1989) The authoring of hypertext documents. In: R. McAleese (ed.) *Hypertext: Theory into Practice*, Oxford: Intellect, 138-147.

McKnight, C., Dillon, A., and Richardson, J. (1990b) *Hypertext in Context.* Cambridge: Cambridge University Press

Medawar, P. (1964) Is the scientific paper a fraud? In: D. Edge (ed.) *Experiment.* London: BBC Publications.

Mills, C.B. and Weldon, L.J. (1985) Reading text from computer screens. Centre for Automation Research, Human-Computer Interaction Laboratory, University of Maryland, Maryland 20742.

Milner, N. (1988) A review of human performance and preference with different input devices to computer systems. In: D. Jones and R. Winder (eds.) *People and Computers IV.* Cambridge: Cambridge University Press. 341-362.

Monk, A., Walsh, P. and Dix, A. (1988) A comparison of hypertext, scrolling, and folding as mechanisms for program browsing. In: D. Jones and R.Winder (eds.) *People and Computers IV.* Cambridge: Cambridge University Press, 421-435.

Muter, P., Latremouille, S.A., Treurniet, W.C. & Beam, P. (1982) Extended reading of continuous text on television screens. *Human Factors*, 24(5), 501-508.

Neal, A. and Darnell, M. (1984) Text editing performance with partial line, partial page and full page displays. *Human Factors,* 26(4), 431-441.

Nelson, T., (1987) *Literary Machines.* Abridged Electronic Version 87.1 San Antonio: Ted Nelson.

Newell, A. and Card, S. (1985) The prospects for psychological science in Human-Computer Interaction. *Human-Computer Interaction*, 1, 209-242.

Nisbett, R. and Wilson, T. (1977) Telling more than we can know: verbal reports on mental processes. *Psychological Review,* 84, 231-259.

Norman D. (1986) Cognitive Engineering. In: D. Norman, and S. Draper (eds.) (1986) *User Centred System Design.* Hillsdale NJ: Lawrence Erlbaum Associates, 31-61.

Norman, D. and Draper, S. (eds.) (1986) *User Centred System Design.* Hillsdale NJ: Lawrence Erlbaum Associates.

Norman, K. and Chen, J. (1988) The effect of tree structure on search in a hierarchical menu selectration system. *Behaviour and Information Technology*, 7(1), 51-65.

Oborne, D. and Holton, D. (1988) Reading from screen versus paper: there is no difference. *International Journal of Man-Machine Studies*, 28(1), 1-9.

Olshavsky, J. (1977) Reading as problem solving: an investigation of strategies. *Reading Research Quarterly*, 4, 654-674.

Oppenheim, A. N. (1966) *Questionnaire Design and Attitude Measurement.* London: Heinemann.

Osborne, A. (1979) *Running Wild: The Next Industrial Revolution.* Berkeley: McGraw-Hill.

Pastoor, S., Schwartz, E. and Beldie, I. P. (1983) The relative suitability of four dot-matrix sizes for text presentation on colour television screens. *Human Factors,* 25(3), 265-272.

Pearce, B. (ed) (1984) *Health Hazards of VDUs?* Chichester: John Wiley and Sons.

Pearson, R.G. and Byars, G.E. (1956) The development and validation of a checklist for measuring subjective fatigue. Report # TR-56-115, San Antonio, TX: USAF School of Aviation Medicine.

Polson, P., Muncher, E., and Kieras, D. (1987) Transfer of skills between inconsistent editors. Tech. Rep. No. 87-10. Boulder: University of Colorado, Institute of Cognitive Science.

Polson, P., Muncher, E. and Engelbeck, G. (1986) A test of a common elements theory of transfer. In M. Mantei and P. Orbeton (eds.) *Proceedings of CHI'86.* New York: Association for Computing Machinery, 78-83.

Popper, K. (1972) *The Logic of Scientific Discovery* (Third Edition). London: Hutchinson.

Pugh, A. (1979) Styles and strategies in adult silent reading. In: P. Kolers, M. Wrolstad and H. Bouma (eds.) *Processing of Visible Language 1.* London: Plenum Press.

Pullinger, D. (1984) Design and Presentation of the CHF journal on the BLEND system. *Visible Language,* 18(2). 171-185.

Radl, G.W. (1980) Experimental investigations for optimal presentation mode and colours of symbols on the CRT screen. In E. Grandjean and E. Vigliani (eds.) *Ergonomic Aspects of Visual Display Terminals.* London: Taylor and Francis, 127-137.

Rasmussen, J. (1986) *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. London: North Holland.

Richardson, J., Dillon, A. and McKnight, C. (1989) The effect of window size on reading and manipulating electronic text. In E. Megaw (ed.) *Contemporary Ergonomics 1989*. London: Taylor and Francis, 474-479.

Richardson, J., Dillon, A., McKnight, C. and Saadat-Samardi, M. (1988) The manipulation of screen presented text: experimental investigation of an interface incorporating a movement grammar. HUSAT Memo #431, Loughborough University of Technology.

Rivlin, C., Lewis, R. and Cooper, R. (1990) *Guidelines for Screen Design*. Oxford: Blackwell Scientific.

Rothkopf, E. Z. (1971) Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behavior*, 10, 608-613.

Rumelhart, D. (1977) Toward an interactive model of reading. In: S. Dornic (ed.) *Attention and PerformanceVI*. Hillsdale NJ: Erlbaum.

Ryle, A. (1976) Some clinical applications of grid technique. In: P. Slater (ed.) *The Measurement of Intrapersonal Space by Grid Technique* (2 vols.). London: John Wiley and Sons.

Samuels, S. and Kamil, M. (1984) Models of the reading process. In: P. Pearson. (ed.) *Handbook of Reading Research*. New York: Longman, 185-224.

Sauter, S., Gottlieb, M., Rohrer, K. and Dodson, V. (1983) The well-being of video display terminal users: an exploratory study. Report No: 210-79-0034. Cincinnati, OH: US Dept. of Health and Human Sciences.

Schumacher, G. and Waller, R. (1985) Testing design alternatives: a comparison of procedures. In: T. Duffy and R. Waller (eds.) *Designing Usable Texts*. Orlando FL: Academic Press, 377-403.

Schwartz, E., Beldie, I. and Pastoor, S. (1983) A comparison of paging and scrolling for changing screen contents by inexperienced users. *Human Factors*, 25(3), 279-282.

Sekey, A. and Tietz, J. (1982) Text display by "saccadic scrolling". *Visible Language*, 17, 62-77.

Shackel, B. (1987) An overview of research on electronic journals. In: G. Salvendy (ed.) *Cognitive Engineering in the Design of Human Computer Interaction and Expert Systems*. Amsterdam: Elsevier,193-206

Sharratt, B. (1987) The incorporation of early interface evaluation into command language grammar specifications. In D.Diaper and R. Winder (eds.) *People and Computers III*. Cambridge: Cambridge University Press, 11-28.

Shaw, M.L.G. (1980) *On Becoming a Personal Scientist*. London: Academic Press.

Shaw, M. L.G. and Gaines, B. (1987) KITTEN: Knowledge initiation and transfer tools for experts and novices. *International Journal of Man-Machine Studies*, 27, 251-280.

Shneiderman, B. (1987) *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. San Francisco: Addison Wesley.

Shneiderman, B. (1984) The future of interactive systems and the emergence of direct manipulation. In: Y. Vassiliou (ed.) *Human Factors and Interactive Computer Systems*. Norwood NJ: Ablex, 1-27.

Shulman, H., Shute, S. and Weissmann, O. (1985) Icons versus names as command designators in text editing. *Proceedings of the International Conference on Cybernetics and Society*. New York: IEEE, 268-272.

Simpson, A. (1989) Navigation in hypertext: design issues. Paper presented at International OnLine Conference '89, London, December.

Simpson, A. (1990) Towards the design of an electronic journal.Unpublished PhD Thesis. Dept. of Human Sciences, Loughborough University of Technology.

Simpson, A. and McKnight, C. (1989) Navigation in hypertext: structural cues and mental maps. In: R. McAleese and C. Green (eds.) *Hypertext: State of the Art.* Oxford: Intellect, 73-83.

Slater, P. (1976) *The Measurement of Intrapersonal Space by Grid Technique (2 vols.)* London: John Wiley and Sons.

Smith, F. (1978) *Reading*. Cambridge: Cambridge University Press.

Stammers, R., George, D. and Carey, M. (1989) An evaluation of abstract and concrete icons for a CAD package. *Contemporary Ergonomics 1989*, London: Taylor and Francis, 416-421.

Starr, S.J. (1984) Effects of video display terminals in a business office. *Human Factors*, 26(3), 347-356.

Suchman, L. (1988) *Plans and Situated Action*. Cambridge: Cambridge University Press.

Switchenko, D. M.(1984) Reading from CRT versus paper: the CRT disadvantage hypothesis re-examined. *Proceedings of Human Factors Society, 28th Annual Meeting*. Santa Monica CA: Human Factors Society, 429-430.

Tinker, M. A. (1963) *Legibility of Print*. Ames, Iowa: Iowa State University Press.

Tinker, M.A. (1958) Recent studies of eye movements in reading. *Psychological Bulletin*, 55, 215-231.

Tolman, E.C. (1948) Cognitive maps in rats and men. *Psychological Review*, 55, 189-208.

Tombaugh, J., Lickorish, A. and Wright, P. (1987) Multi-window displays for readers of lengthy texts. *International Journal of Man-Machine Studies*, 26, 597-616.

Trigg, R. and Suchman, L. (1989) Collaborative Writing in NoteCards. In: R. McAleese (ed.) *Hypertext: Theory into Practice*. Norwood NJ: Ablex, 45-61.

Tuck, B., McKnight, C., Hayet, M. and Archer, D. (1990) *Project Quartet.* Library and Information Research Report 76, London: The British Library.

Venezky, R.L. (1984) The history of reading research. In: P. Pearson (ed.) *Handbook of Reading Research.*New York: Longman, 3-38.

Ventura, C. (1988) Why switch from paper to electronic manuals? *Proceedings of the ACM Conference on Document Processing Systems*, Santa Fe NM: Association for Computing Machinery, 111-116.

Waern, Y. & Rollenhagen, C. (1983) Reading text from visual display units (VDUs). *International Journal of Man-Machine Studies,* 18, 441-465.

Waller, R. (1984) Designing government forms: a case study.*Information Design Journal,* 4, 36-57.

Waller, R. (1986) What electronic books will have to be better than. *Information Design Journal* , 5, 72-75.

Waller, R. (1987) The typographic contribution to language: towards a model of typographic genres and their underlying structures. Unpublished PhD Thesis, Dept. of Typography and Graphic Communication, University of Reading.

Whalley, P. and Fleming, R. (1975) An experiment with a simple recorder of reading behaviour. *Programmed Learning and Educational Technology,* 12, 120-124.

Whitefield, A. (1989) Constructing appropriate models of computer users: the case of engineering designers. In J. Long and A.Whitefield.(eds.) *Cognitve Ergonomics and Human Computer Interaction.* Cambridge: Cambridge University Press, 66-94.

Wickens, C. (1984) *Engineering Psychology and Human Performance.* Columbus: Charles Merrill.

Wilkinson , R.T. and Robinshaw, H.M. (1987) Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour and Information Technology,* 6(2), 125-133.

Wilson, M.D., Barnard, P.J. and Maclean, A. (1986) Task analysis in human-computer interaction. Hursley Human Factors Laboratory Report HF122, Hursley Park, Winchester

Winograd, T. and Flores, F. (1988) *Understanding Computers and Cognition.* Reading MA: Addison Wesley.

Wittgenstein, L. (1953) *Philosophical Investigations*. New York: MacMillan.

Wright, P. (1980) Textual literacy: an outline sketch of psychological research on reading and writing. In: P. Kolers, M. Wrolstad and H. Bouma (eds.) *Processing of Visible Language 2*. London: Plenum Press.

Wright, P. and Lickorish, A. (1983) Proof-reading texts on screen and paper. *Behaviour and Information Technology,* 2(3), 227-235.

Wright, P. and Lickorish, A. (1984) Ease of annotation in proof-reading tasks. *Behaviour and Information Technology,* 3(3), 185-194.

Wright, P. and Lickorish, A. (1988) Colour cues as location aids in lengthy texts on screen and paper. *Behaviour and Information Technology,* 7(1), 11-30.

Zechmeister, E. and McKillip, J. (1972) Recall of place on a page. *Journal of Educational Psychology*, 63, 446-453.

Zechmeister, E., McKillip, J., Pasko, S. and Bespalec, D. (1975) Visual memory for place on the page. *Journal of General Psychology*, 92, 43-52.

Appendices

Appendix A: The ADONIS Document Delivery Workstation Evaluation Report

Appendix B: Complete transcription of verbal protocol for one subject in utility experiment

The ADONIS Document Delivery Workstation: an Interface Evaluation

Andrew Dillon
HUSAT Research Centre
Elms Grove
Loughborough
LEICS. LE11 1RG

PQ/LUT 26
July 1988

Abstract

This document presents a usability evaluation of the ADONIS Document Delivery Workstation. This system was designed for use in various document supply centres by library staff familiar with the principles of document delivery. However, any successful implementation of such a system is likely to have applications in the public domain, and the present evaluation therefore considers the ADONIS system from both points of view. Human factors guidelines pertaining to the interface are presented and the match between these and the ADONIS workstation is critically assessed in the light of investigations carried out at the British Library Document Supply Centre in Boston Spa and at HUSAT. Recommendations for improving the interface are presented in the final section of the report.

1. Introduction

The ADONIS Document Delivery workstation is designed to facilitate searching, viewing and printing of CD-ROM stored articles. ADONIS was developed in order to investigate the extent to which information technology could reduce the costs of labour intensive photocopying procedures in the document supply domain and to increase copyright control over published material (Campbell and Stern, 1987). Biomedical journals were selected for the trial service on the basis of journal usage studies (Clarke, 1981).

The workstation consists of an IBM PC/AT or compatible (the present study was carried out on an NEC Powermate 2), a Hitachi CD-ROM drive, a high resolution A4-size monitor and a Ricoh laser printer. Basic interaction with the system requires the user to enter data via a form filling screen and select options from a menu of commands. Loading of CD-ROMs is also necessary for the display of full text and printing. Thus interaction is constrained to a relatively simple sequence of tasks.

The intended users of the ADONIS system are document supply personnel who come to the workstation with specific article requests. Since they employ the technology as part of their normal work duties they will be referred to here as the dedicated users. However any successful implementation of such a workstation is likely to have applications with end-users in libraries who will undoubtedly find full text document delivery of on-line material advantageous. Such users will be referred to here as the casual users.

In order to appreciate how well dedicated users could employ this technology a visit was arranged the the British Library Document Supply Centre (BLDSC) at Boston Spa to discuss the system with four such users and observe their routine interactions. The emphasis for this part of the evaluation was on how well they used the system, what difficulties they had and how they would like the system to be improved. To appreciate the usability of the system from the point of view of casual users, ten ADONIS-naïve users were asked to perform three tasks in guided interactions with the system. An evaluator observed their use of the system and recorded their impressions of the interface. The emphasis here was on their ability to interact successfully without training or formal introductions to the system, as is typical of systems aimed at casual users in the public domain.

As mentioned earlier, user interaction with the system is limited to form filling for searching and printing, use of function keys for mode selection, page up/page down keypresses for display manipulation, and loading CDs. Form filling is one of the more common means of facilitating casual user-system interaction and Shneiderman (1987) has proposed a set of guidelines for the design of these interfaces. On-screen manipulation of text has received relatively little attention to date though work is being carried out at HUSAT by the Project QUARTET team on this subject. The following section will outline the human factors principles that have emerged in these areas, and will detail how well ADONIS fares on these criteria, incorporating the results of the evaluations. General user ratings and comments of the system are provided. The final section contains a list of recommendations for improving the interface for the respective user groups.

2. Human Factors Guidelines and the ADONIS System.

In terms of the human factors literature, form filling and menu selection have been well researched as suitable means of interaction, and guidelines for successful design have been proposed. Navigation of text, however, is a relatively new problem and research interest in this domain is in its infancy. This section will present the guidelines relevant

to ADONIS and comment on the match between these and the present system.

2.1. Form filling

Shneiderman (1987) lists 9 guidelines for the design of form-filling interfaces. These represent a good condensation of the findings of human factors research in the area.

2.1.1. *Meaningful title*

It should be clear from the display what is the subject matter of the form. The topic should be clearly identified and computer jargon should be avoided. This is not too important here given that ADONIS is a bespoke database for retrieving journal articles only and will rarely be used by individuals with no realisation of its uses.

2.1.2. *Comprehensible instructions*

This guideline is particularly relevant to systems that will be utilised by occasional or infrequent users who cannot be expected to retain knowledge of system operations between interactions and therefore may need to be briefly re-introduced to the system each time. Instructions should be brief and avoid pronouns or references to the user. The style of instruction should be consistent across screens or forms.

The instructions provided with the ADONIS system are minimal. When switched on, the system provides the user with a menu. Once the document specification form has been arrived at, the user is provided with no instructions on how to proceed; rather, they receive a prompt for the ADONIS number which initially fails to mean anything. This is unlikely to be a cause of much difficulty to relatively frequent users, indeed dedicated users had few problems here, but it seemed to confuse the casual users.

2.1.3. *Logical grouping and sequencing of fields*

Related fields should be grouped together and sequencing should reflect common patterns. The idea behind this guideline is that the required information will inevitably have a structural logic to it that will determine the user's storage and recall. Prompting for information in that order will increase the chances of a successful interaction with the system. Shneiderman provides the obvious example of address filling where town or city should be grouped with county and post code.

The actual ordering and layout of fields on ADONIS is poor. The document specification form is laid out in the following sequence:

ADONIS number / ISSN number / journal title / year / author / article title / volume / part / page number.

This is an odd sequence that owes more to the programming of the database than users' tendencies to structure their description of articles.

A survey of over 30 individuals at HUSAT showed that references are typically described in the sequence:

author / article title / journal title / year.

While some variance exists (e.g., 'author / year / article title / journal title' is also very common), details such as volume, page-numbers, part or ISSN number are seemingly rarely used to describe a reference.[1]

This had observable effects on casual users who displayed a tendency to enter article

titles in the journal field. Upon realising the error these users explained their actions in terms of having seen the word 'title' in the prompt and assuming that it must refer to the title of the article. In other words, users structure their descriptions of references in a way that disposes them to think of the article title before the name of the journal.

The strategy adopted by the dedicated users involves input of "journal / volume / part / page' details only. This works effectively most of the time and reduces the number of keypresses required. Should it fail, they proceed to supplement the search information with an author name or a word or two of the article title. These users claim never to input more than one author or the complete article title as much of it is redundant information. The ADONIS number is rarely available and these users commented that it is unlikely to provide an improvement as copying a 16 digit code is so prone to error on the part of the document requester and/or the operator that supplementary information will still need to be entered.

These facts would suggest that the sequencing of fields should be altered for both users to support their distinctive search strategies. As it is, the sequencing fails to support either user type.

2.1.4. *Visually appealing layout*

It is important that spacing and field alignment are given some thought in order to provide a clear and uniform distribution of fields throughout the form. Alignment is considered to create a sense of order and comprehensibility. In particular, good layout directs the user's attention to areas of interest that require input or response.

It must be said that layout is one of the poorest aspects of the ADONIS system. Given the size of the screen most casual users felt that the main window was poorly positioned at the bottom. All users felt that the hatching design of the fields on the search form was poor; it tended to dominate the image and rendered fast location of the cursor difficult.

Alignment of labels and fields is inconsistent; sometimes label is above field, e.g., article title, and other times it is to the left of it, as in the ADONIS number field. This creates a disorderly and untidy effect. Furthermore, the fact that all user input is automatically presented on screen in identical font and case to the input prompts can lead to difficulties in discrimination of user and computer generated text.

The preponderant use of upper case lettering throughout the interface is undesirable. Upper case letters lack ascenders and descenders and thus reduce word shape resulting in poorer readability of the presented text.

Dedicated users also criticised the print request form for being badly laid out, and even though an automatic addressing system has now been linked up there are difficulties inputting post codes, and users must still manually input information about the delivery route code.

2.1.5. *Familiar field labels*

The basic point here is that unusual terminology should be avoided, e.g., "identifier" instead of "name" or "domicile" instead of "address". ADONIS offers sensible field labels and users seemed to have little difficulty with these.

2.1.6. *Consistent terminology and abbreviations*

The terms used to describe fields or system actions should not alter across forms or modes. In other words requests for "author" should not later become requests for "name", and abbreviations that are acceptable in one field or form should be acceptable

in all.

Inconsistent terminology was not perceived to be a major problem with this system. ADONIS offers the means to abbreviate input in any search field and allows wildcarding of single letters or complete strings. These procedures seem to work effectively though they reduce the search speed considerably. Dedicated users seem to experience some difficulties with the abbreviation facilities but attribute this to lack of practice (a sentiment which highlights their uncritical approach but does not detract from the fact that they had difficulties). In effect they do not use the wildcard facilities, preferring to abbreviate by shortening words as deemed suitable. Casual users adopted similar abbreviation strategies though were consistently more inclined to enter complete information, perhaps highlighting their unfamiliarity with the system and thus, unfortunately, increasing their chances of inputting a mistake.

2.1.7. *Error correction for characters and fields*

The ability to move freely about the input fields to edit typed entries is an essential usability criterion. Ideally back/forward spacing, field clearing, insertion and overtyping should all be provided to facilitate this process. Where input must be of a particular type, e.g., digits only, it is better if the system informs the user immediately should this be violated.

Editing within fields is carried out by use of backspacing, overtyping, insertion or deleting — all available as special keys. Users familiar with keyboards should not have difficulties with these. Complete deletion of entries in all fields is possible with the "abort" function key which is displayed as an option in the menu at the bottom of the screen. However, clearing of individual fields does not appear to be an option on this system. Some casual users initially struggled with the editing facilities but it is likely that such difficulties are quickly overcome with further use or increased exposure to other systems.

The ADONIS number field actually allows illegal input and will initiate a search on its basis. For example, inputting a string of six alpha characters will result in a user waiting while a search is initiated and the system subsequently responds that no entries have been found. Such obviously incorrect input should be flagged immediately by the system to avoid such occurrences.

2.1.8. *Visual templates for common fields*

This guideline refers to the provision of visible space and boundaries for data entry fields so that the user is given an indication of the size and, perhaps, ordering of the input required. Where precise calendar dates are required it may help to offer a field partitioned into sections for day / month / year, or for telephone numbers to offer a bracketed section for the area code.

The fields in the ADONIS form are of varying sizes to accommodate the details required. The year field offers four spaces, and the ADONIS number field offers sixteen, both clearly appearing filled when the correct information is entered. The title fields for journals and articles obviously cannot be so specific and thus offer window-wide fields which allow maximum input of data. Unfortunately the title field, albeit large, can still prove insufficient for complete titles of documents and this is considered troublesome by casual users who felt that in such cases the screen should adjust to provide space for further input. The actual title template on the system at the BLDSC has been shortened due to comments from users to the effect that they never enter a full title. However it still fills three-quarters of the screen width and the intention of this 'improvement' is difficult to perceive.

The part field can cause difficulties when journals release special issues with two or

more parts in one (e.g., Vol. 1, Parts 1 & 2). The part field in ADONIS will not take hyphens or spaces, thus to specify such an issue the user must input the two digits as one number (i.e., parts 1 & 2 become part 12). No reference to this fact could be found in the manual. It was discovered by one dedicated user after trial and error.

The page numbers field is also problematic. Ideally, the user need only input the page number at which the article begins. However, some of the casual users attempted to input the complete details (e.g., 111-123). The available field will not accept such large input and furthermore will not allow the input of spaces or hyphens. Again no reference to this limit on page number specification could be found in the manual. Users must infer it for themselves.

2.1.9. *Help facilities*

The user should always be able to access relevant information about suitable input or system activities. Ideally this should be contextually embedded in the interface so that the user receives relevant information to the task in hand when help is evoked. This extends beyond the requirements of form-filling screens to help facilities in all modes. Unfortunately the help in ADONIS is weak and casual users commented that it told them little they could not deduce themselves. One casual user felt that worked examples would prove a better way of packaging information in the help facilities. The dedicated users stated that they did not use the help facilities but proceeded by trial and error until they found a strategy that worked. This can be interpreted as a failure on the part of the help facilities to offer users obviously useful information.

2.1.10. *Summary*

Nine guidelines for the design of usable form filling screens have been presented. For dedicated users the ADONIS system appears satisfactory on five of these. However, the sequencing of fields, the visual presentation and layout of the form, the field templates and the help facilities could be improved. For casual users the interface is noticeably poorer. They expressed dissatisfaction with five of the interface qualities raised in this section. Sequencing of fields, and visual layout are also issues for these users but lack of suitable help, poor instructions and weak template design all caused observable difficulties to this user group.

2.2. Menu selection

Menu based interactions are by far the most popular means of providing access to a system. They are particularly useful for casual or non-specialist users as they support recognition rather than recall of acceptable input, can overcome the use of technical command languages and can be successfully used with little or no training. Research on menu interfaces suggests that size of menu, particularly the depth and breadth trade-off, organisation and categorisation of the available options and navigation facilities are important design issues for successful menu-based systems.

With reference to the present system the following guidelines are pertinent:

2.2.1. *Manageable size*

Menus can be described in terms of the number of levels (depth) and the number of items per level (breadth) they contain. Research indicates that generally, users perform faster and more accurately on menu structures that are broad and shallow rather than deep and narrow (Snowberry *et al.* 1983; Kiger, 1984). However, it is important to realise the limits of such claims and appreciate the contextual effects of subject matter, terminology, user group, etc., that all influence the extent to which certain structures are preferable.

In the present situation menu size is constrained by the limited functionality of the system and thus trade-offs between depth and breadth can be handled sensibly. Typically, menu depths of 3 levels and breadths of 4 or 5 items are the ADONIS size and users experienced few problems with size-related issues.

2.2.2. *Meaningful organisation*

It is important that items are grouped into menus in a meaningful fashion, ideally reflecting an established or well understood categorisation structure. For example, a menu based system for providing geographic information should provide menus structured in terms of country on the top level, counties or states on the next level and specific places on the third level rather than several menus offering a complete alphabetic listing of all specific place names available on the system. Where it is difficult to identify such a meaningful structure, task analysis of the intended users can provide insight into how these individuals organise information and the menus should be constructed accordingly.

At the highest level, ADONIS structures menus according to task activities, i.e., does the user wish to search, print, or carry out a maintenance function? Once a selection is made the next level offers several functions specific to that selection and so forth. The organisation of the available options appears sensible, however it is noteworthy that several casual users remarked on the distinction between printing and searching for an article at the top level. Being offered the print option at this point suggested to them that they should take this route to obtain hard copy of an article and the search route if they only wanted to view it on screen. Similarly, dedicated users do not use option 4 of this menu (report generation) and thus felt it should not be offered to them.

2.2.3. *Ease of selection*

Many of the advantages of menu based interfaces are lost if users have to spend time and effort indicating their choice of option to the system by precise typing of the name. It is important therefore that users can specify a choice with the minimum of input and delay. Thus single number or first letter specifications are better than full typing of selection; avoiding the need to hit the "return" key after selection can speed up interaction, as can type-ahead facilities which, if implemented consistently, allow frequent users to navigate quickly through familiar paths.

The ADONIS system has opted for a mix of both number input and function key mechanisms for menu selection. Thus at the start of interaction, at the mode selection level, users must select a number from 1 to 4 to nominate the task they wish to perform or hit a function key to get help or to quit. Such mixed mechanisms are acceptable as they attempt to distinguish between mode selection (task to be performed) and system options (help, quit, initiate procedure, etc.). However there is an apparent inconsistency in the distinction between the function keys and the number menus. Typically the latter are displayed in the interaction window presented to the user while function key options line the bottom of the screen. However, if the user selects print mode from the top menu the function key options are presented in the window and the usual line of highlighted options at the bottom of the screen disappears. Furthermore six of the 10 casual users initially failed to appreciate the distinction between function and numeric keys which caused several of them difficulty in getting started. These are learning problems that may deter casual users from further unsupervised interaction.

2.2.4. *Clear terminology*

The options presented to the user in a menu should be self-explanatory. A casual user should be able to discriminate between options and have a reasonable degree of confidence about what the available options will accomplish. Obviously this requires careful selection of terms and the avoidance of jargon, unnecessary technical language

or slang. Ideally, the language of the intended user population should be maintained with the computerisation of the task.

A number of the terms in the ADONIS menus were perceived as less than clear, particularly by the casual users. For example "exit" is a classically ambiguous term to inexperienced users who often think it means leaving the application rather than the current point of interaction. "Abort" is a universally unpopular word amongst casual users and should thus be avoided. In the context in which it is used in this system, i.e., to delete a full form and empty the field contents, the word "Clear" is a more sensible alternative. The term "display" was felt by many to be misleading as it does not, as is implied, display the contents of an article but rather presents the full reference information of the article. The extent to which these and similar difficulties will be overcome by practice cannot be overlooked, though for casual users they all add to the difficulties of operating ADONIS effectively.

2.2.5. *Summary*

The ADONIS menus are acceptable in terms of size, organisation and ease of selection, especially for the dedicated users. However casual users expressed some difficulties with the terminology of, and distinctions between, several options. Problems such as these are likely to deter such users from persisting with the system unsupervised. Minor modifications would probably resolve these difficulties.

2.3 Text Manipulation

As stated earlier, formal guidelines on interfaces for the manipulation of text have yet to be developed. However on the basis of the traditional human factors literature and our own research experience it is possible to provide generic advice on how such systems should be designed. Ideally, facilities to move freely about the text with minimum effort or delay should be available to the user. These may include such features as going to a specific page or place in the text such as "Figure 1" or "the References"; continuous presentation of navigational information; signposting of current position; ease of exit both to one's previous position and out of the text completely. Scrolling through the text should not cause the screen image to jump, and users should not be faced with delays between presentation of text sections such that unnecessary reliance is placed on their memory. It is likely that considerations such as these will prove important determinants of the success of such systems.

We are aware that the ADONIS system stores articles as bit-mapped images, not as straight ASCII files. Therefore the speed at which images can be presented and the facilities available for manipulation of text are severely restricted. Furthermore the system was not designed for the display and manipulation of text and dedicated users at the DSC have little use for the display facilities. These facts must be borne in mind during the evaluation of this aspect of the system.

2.3.1. *Navigational information*

In the case of ADONIS, when an article is displayed on screen all information about navigation and manipulation disappears and the user is left with no instructions. This is very unsatisfactory. The user is supposed to remember that F9 will toggle the information window on/off and that page-up/down keys will manipulate the document. As it happens, pressing virtually any key will result in an error that fortunately causes the search details and instruction menu to reappear over the bottom half of the document. While this may appear simple enough (though somewhat inelegant) the page-up/down keys are duplicated on certain keyboards. Users have been observed in difficulty at this point wondering why the numeric pad page up/down keys they are pressing are not effective. It would improve the interface if manipulation commands were permanently displayed in a menu at the bottom of the screen and ambiguities

regarding the special keys were removed.

2.3.2. *Range of facilities*

Navigational facilities are also limited. The user may proceed through the document page by page, but it is not possible to jump or 'goto' specific places. All casual users remarked that such a facility was essential. Speed of navigation is also very slow and the time taken to present a new page severely restricts browsing of articles. All casual users disliked the presentation style, preferring either the ability to scroll across pages or total page presentation instantly rather than the slow fill-from-the-top style of this system.

2.3.3. *Summary*

Manipulation of text is difficult on the present system. Even though there are understandable reasons for this, casual users are likely to demand faster and more flexible facilities. At present, essential navigational information is too easily lost and the inability to jump to specific places or pages is seen as severely restrictive.

2.4. More general usability issues

Apart from formal guidelines of certain interface designs such as forms and menus, there are a number of general characteristics of a system that will determine its usability. These are such qualities as levels of feedback, language and tone of messages, error handling facilities, etc. With reference to ADONIS the following aspects appear relevant:

2.4.1. *Feedback*

Provision of feedback to the user is a fundamental human factors principle for interface design. Feedback informs users when input is expected and of the success or failure of their actions. A standard example of such feedback is a message informing the user that a delete operation has been successfully completed. Generally ADONIS was perceived to provide little explicit feedback though this caused more difficulties for casual than dedicated users.

Particularly commented upon was the feedback provided when a search is being carried out. The user is informed that the system is "searching" and provided with an option to interrupt this activity. However search times of up to seven minutes have been noted and such feedback is insufficient for users waiting on a response. Most casual users remarked that the red light denoting hard disk activity provided more information on what was happening than the interface. Ideally users would like some indication of how long a search may take or some form of concurrent feedback that informed them that the system was not looping or 'hung'.

An example of inappropriate feedback occurs when a search on authors is carried out. If 'hits' occur, these are presented to the user as a list of titles and the user has no way of knowing whether these articles are all by the same author, by authors with similar names or by the required author and several colleagues. While users may resolve this for output of only a few titles by displaying document details for each title, for searches which result in many more titles this is tedious and time-consuming, particularly, as dedicated users remarked, since there is no apparent order to this output. Interestingly, searching on titles also causes feedback in terms of titles.

The dedicated users pointed out that when batch printing, the system fails to discriminate between signalling that a print out has been completed and that a new CD is required. Both result in a single "bleep" from the computer. Thus if working elsewhere while the system is batch printing one must guess which signal is which and

regularly check the status of the machine. This removes a major advantage of the system, i.e., allowing users to concentrate on other work while articles are being output.

2.4.2. *Language of System/Error messages*

The language of the system messages, prompts, etc., is another crucial human factors issue. To some extent this issue has been raised in discussions on terminology and menu options and these will not be discussed further here. However the system also provides messages to the user through prompts for disks or in response to user errors.

The ADONIS software manifests obvious consideration of system language issues at certain points, for example the prompts for users to enter particular disks or to check that the disk drawer is closed are clear and concise. However at the other extreme, error messages exist which are extremely hostile and uninformative. One example occurs if the user selects option 4 (Report generation) from the top menu when the printer is not switched on. This causes a system response of:

> Fatal error in <report_main> code 0001 ffff

From a human factors perspective such messages are useless, offering no indication why an error occurred or what a user should do to rectify the situation — essential features of any diagnostic message. Errors such as incorrect keypresses often result in a sharp tone from the machine which certainly informs the user of an error but is also less than "user-friendly"

2.4.3. *Input formality*

Input formality refers to the extent to which user input must satisfy specific and rigid parameters for successful interaction. A highly formal system will tolerate no ambiguity of input, and spelling mistakes or keying errors will result in totally inaccurate output. Obviously, such systems are likely to cause non-specialist or infrequent users considerable difficulties.

Whereas ADONIS allows abbreviation of search criteria (see Section 2.1.6) it is extremely intolerant of simple inclusions or omissions of punctuation. For example, the presence of a comma between an author's surname and initial when searching on a name only has been found to produce results different from an identical search without the comma. Thus the system may inform the user that there are no articles for example, by Smith, D.M. even though a search on Smith or Smith D.M. (without the comma) may reveal several. Similar effects have been noted for the presence or absence of a hyphen in article titles. Simple spelling mistakes have the same consequences. This is extremely irritating and can lead users to draw totally incorrect conclusions from the database.

2.4.4. *Help facilities*

Help facilities in form filling have already been mentioned in Section 2.1.9. The facilities available in other modes are similarly accessed by pressing function key 1, whereupon a window containing the help information is presented on screen. As mentioned previously, dedicated users claimed to have no use for these facilities; therefore the following points relate to the casual users' perceptions only.

Several of the casual users attempted to access help upon first exposure to the system, i.e., at the mode selection menu. Unfortunately, even though help is offered here the

facilities only contain a message to the effect that they have not been actually implemented at this point. At other stages users complained of the lack of information contained in help, with only single line explanations of commands being offered. Obviously the facility to query-in-depth by asking for further information would be desirable. Layout of the help windows was also criticised for containing information about function keys that appeared selectable at that point when in fact users had to exit help before selecting such options. In conclusion, the help facilities were considered of little use by both type of users.

2.4.5. Ease of exit

Users should be able to interrupt their interactions at any point and backtrack to previous stages. Ideally a single command or key should accomplish this action in all situations. In a system such as ADONIS such a facility would allow users to retrace their steps from the deepest level (e.g.document display) to the highest (e.g.task selection) with a simple keystroke per level.

To some extent the provision of a dedicated function key (F10) which acts as a step back from the current to previous level satisfies this criterion. However, if users exit from article display mode by pressing F10 they are returned to an empty search form, rather than the form containing the criteria they searched on thus losing all input information pertinent to that article. This is particularly frustrating if a search results in multiple 'hits' and users wish to refine their search criteria. Rather than being able to enter additional information on the search form they must start the whole procedure again. Such inconsistency of system response should be corrected.

2.4.6. *Summary*

ADONIS fails to satisfy several instances of the general usability criteria outlined in this section. The step back facility is inconsistent and can be a source of frustration. Provision of feedback and the help facilities were criticised by both dedicated and casual users. System language and input formality were more of a problem for the casual users.

2.5. General user ratings of ADONIS

Obviously the users tested in this evaluation form two quite distinct user groups. The dedicated users at the DSC are the group for whom the system was designed and therefore their comments are directly relevant to the task of evaluation. The users studied at HUSAT are representative of the type of user who would interact with this system were it available to end-users in libraries. While their comments may not be seen as equivalently relevant, it is conceivable that ADONIS will be made available in this way, thus such responses are pertinent to any future application of this system.

2.5.1. *Dedicated users*

There was a general consensus amongst these users of the value of the ADONIS system. They could perceive the advantages of computerised storage and retrieval and felt that the quality of service they were able to offer with this system was good. In terms of the interface, the main body of opinion was that they were all able to use it reasonably well thus 'it must be satisfactory'! Lack of experience with other computer systems made critical comparisons impossible. However they specifically felt that the search form could be better laid out to suit their search strategy (see Section 2.1.3) and they all wanted the ability to print and search simultaneously.

The large screen was seen as wasteful as they rarely used the display facilities except, as reported earlier, to check the article if a print error occurred. Size of the system rendered positioning difficult and the layout of the system at the DSC site was

ergonomically poor. Research (e.g. Grandjean 1984) indicates that the screen centre should be between 90-115 cm. above the floor facilitating a down viewing angle (eye level to screen centre) of approximately -9°. The ADONIS workstation is significantly outside this range. Placing the screen directly on the desk instead of on the main body of the computer would improve this but only by greatly increasing the system's footprint. Deskspace at BLDSC is becoming increasingly constrained as the number of disks increases and printouts are currently stacked on a spare chair. Furthermore, background lighting reflects strongly off the screen, a problem that could be alleviated by reversing the screen polarity to black on white. Doing so would exploit the obvious quality of the screen by enhancing screen image.

None of these users received formal training. Typically they had been introduced to the system by another user and been expected to do likewise themselves. The manual with which they are provided tends to be avoided as they consider it too technical. Discussion of this point revealed that they had not actually read the manual but had browsed the first few pages which deal with system installation. They seemed surprised when it was demonstrated that the contents of chapters 5 and 6 were more relevant to their needs. This suggests the need for better separation of these issues in the manual, and formatting of contents to suit the users. Given that users are notorious for avoiding documentation, it is important that this situation is not exacerbated by poor manual design. Provision of relevant help facilities may improve the situation.

In conclusion, these users find the system tolerable but would not want to work full-time on it. They found that several problems with both the hardware and software had emerged since its introduction which gave them the impression that the system was "unreliable". This lack of reliability was seen as the major drawback, necessitating constant supervision of output from ADONIS rather than freeing them to do other work. They claimed that using the system was actually slowing them down rather than speeding them up but realised that once improvements in reliability were achieved this may change. Their general lack of experience with information technology rendered them very uncritical and willing to tolerate inadequacies in the system that less naïve users would find unacceptable.

2.5.2. *Casual users*

These users were much more critical of the system. Apart from the issues raised in previous sections, common criticisms related to the rather "archaic" style of the ADONIS interface. Several users remarked that it looked "old-fashioned" or felt "like something I used years ago", suggesting that a more contemporary interface style incorporating tab keys for moving, highlighting bars instead of flashing cursors, etc., would be preferable.

All casual users felt that the system was far too slow for the types of searches they would normally perform, e.g., attempting to locate articles on the basis of an author and year or author and title. Lack of speed was seen as the one drawback that would severely restrict their usage of such a system were it publicly available. Inability to search on keywords was seen as a further disadvantage.

These users also felt that the size of the system was quite imposing and made it difficult to position oneself comfortably. This is important for any interactive situation but particularly for one where the user may be expected to spend long periods of time sitting at the terminal perhaps reading an article.

The provision of rapid access to CD-based material was seen as a positive benefit that would be readily exploited if easily available. However the interface manifest on the current ADONIS system was seen as decreasing any potential benefit by making access overly difficult to casual users. Obviously the bit-mapped images employed with this system limit the speed and facilities available, but it seems as if casual users

will require more usable interfaces and demand the features outlined in Section 2.3. before willingly accepting such technology.

In an attempt to understand the likelihood of potential users actually reading text on screen, they were asked to comment on the readability of the displayed document independently of the manipulation facilities offered. The responses were as follows:

Fine/would read and study as normal:	2
O.K. / Scan only and get print out:	6
Awful, would not use it:	2

In other words such technology is seen at this time primarily as a searching and retrieval tool; users prefer hard copies of documents for their own use. This is not surprising, but it highlights the waste of the good quality, high resolution screen used in the ADONIS workstation.

In conclusion, the interface was received so badly by casual users that it must be rated as unsuitable for their needs. Without considerable re-design it is unlikely that this system would be utilised effectively in the public domain.

3. Recommendations

Given the significant differences between the needs and tasks of the two user groups investigated in this evaluation, recommendations for improving the ADONIS interface will be divided into those that are common, those for dedicated and those for casual users.

3.1. Common recommendations

3.1.1. Hatching of fields in all form filling screens should be removed. An outline of the extent of the field should remain to give an indication of the format of required input.

3.1.2. User input should appear on screen in a different style from system prompts in order to aid distinction.

3.1.3. Alignment and layout of fields in form filling screens should be improved to present a neater and more consistent image.

3.1.4. Editing facilities should be improved so that the ability to clear the contents of a single field with one keypress is available. Keyboards with special or duplicated keys should reproduce, but not replace, the effects of other keys.

3.1.5. Help facilities should be available at all stages of interaction and should contain information relevant to the task in hand. Facilities that allow users to access information at increasing levels of detail are desirable.

3.1.6. Users should not be offered menu options that are of no relevance to them or are not valid selections at a particular stage.

3.1.7. The system should be more tolerant of input in certain situations so that inclusion for example of a comma where the system expects a blank does not produce an incorrect search of several minutes duration. If such tolerance cannot be coded then the system should immediately refuse to accept the comma as input.

3.1.8. Field templates should support the user in specifying the required information,

therefore both the part and page number templates should be redesigned to conform to the demands of the task.

3.1.9. The step-back facility should allow users to return to the search form that contains their search criteria.

3.1.10. Greater use of lower case lettering should be employed, particularly when displaying lists of journal and article titles, to aid readability.

3.2. Dedicated users

3.2.1. The sequencing of fields in the document search form should be organised to suit the search strategy of these users. As mentioned in Section 2.1.3. this appears to be Journal title/ volume/ part / page. The remaining fields should be incorporated in the likely order of their use for further specification.

3.2.2. The use of a unique identifier for each article in the form of an ADONIS number should be reconsidered. Dedicated users expressed strong reservations on the likelihood of success for a system that requires at least two different people (requester and searcher) to reproduce a 16 digit sequence accurately. If this system is to be retained it is worth considering an alternative coding format based on more intuitive grounds, e.g., the initials of the journal title, plus the volume, part and page details.

3.2.3. More explicit feedback is required to distinguish between a system message to change a CD or to inform the user that a successful printout has occurred.

3.2.4. The print request form should be modified to suit the addressing system of the BLDSC.

3.3. Casual users

3.3.1. Comprehensible instructions should be provided at the outset to aid users who have little or no experience of this system. These should contain information dealing with what the system can and cannot do, how to operate it and who to contact in the event of difficulties.

3.3.2. As with dedicated users the sequencing of fields in the document search form should be arranged to suit casual users' observed search strategy of author/ article title/ journal title / year, with the other fields available for further specification.

3.3.3. The terminology of the options available through the menus should be improved to avoid jargon or ambiguous words and to reflect more clearly the actions of the available options.

3.3.4. If the ability to display text on screen is to be exploited more fully, the range of facilities for manipulating text and the quality of navigational information available need to be improved drastically. Though the use of a bit-mapped image of the text reduces the opportunity for wholesale changes in these facilities, it should be possible to allow forward and backward movement to specific pages in these articles and to increase the navigational information on display.

3.3.5. Explicit feedback should be provided for all user input. Particular instances that could be improved are feedback when the system is searching for an article and when several articles result from a general search.

3.3.6. The language of all error messages should be user-oriented and free from unnecessary numeric codes. Ideally such messages should indicate acceptable input.

References

Campbell, R. and Stern, B. (1987) ADONIS — A new approach to document delivery. *Microcomputers for Information Management,* June.

Clarke, A. (1981) The use of serials at the British Library Lending Division. *Interlending Review,* **9**, 111-117.

Grandjean, E. (1984) Postures and the design of VDT workstations. *Behaviour and Information Technology,* **3**(4), 301-312.

Kiger, J. (1984) The depth-breadth trade-off in the design of menu-driven user interfaces, *International Journal of Man-Machine Studies,* **20**, 201-213.

Snowberry, K., Parkinson, S. and Sisson, N. (1983) Computer display menus. *Ergonomics,* **26**(7), 699-712.

Shneiderman, B. (1987) *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison-Wesley: Reading, MA.

Tasks

The following three tasks were employed to make casual users interact with the system.

1. Obtain a print-out of the following article:

Knuckle pads in live-chicken hangars, by Richards (1987).

2. Display on-screen the reference section of the following article

Baker, G.H.B. (1987) Invited review: psychological factors and immunity. *Journal of Psychosomatic Research,* 31,1, pp1-10.
Adonis no: 0022399987000019
ISSN : 0022-3999

3. How many articles by Williamson are in the database?
Display the third article and then print it.

---

[1] The terms author, article title, journal, date, page numbers, volume and part were presented on a sheet of paper and researchers were asked to rank these form one to seven in order of perceived relevance in specifying a reference. Mean rankings were then calculated with low scores indicating high relevance.

Example Protocol for Subject in Validity Experiment

TIME	COMMENT	ACTION
0.00		Reads question 1
0.11	I'm going to the Index to see if there's anything on taste	Scrolls
0.17	No...Contents	Reads Contents
0.24	No....I've a feeling Introduction covers the taste of wine...I'll check that.	Scrolls down to Introduction
0.31		Scrolls further and reads
0.45	...about the colour?	
0.48		Scrolls further and reads
0.53		Scrolls further and reads
0.58		Scrolls further and reads
1.03		Scrolls further and reads
1.06		Has reached new section
1.08		Still reading
1.17	Right I think sweetness is one...	Writes down "Sweetness"
1.20		Scrolls on and reads
1.31		Reaches new section
1.41	I think it's Sweetness and Body, just these two.....	Scanning text, writes down "Body"
1.52	I'll just check to see if there's anything later on.....	Scrolling and reading further
1.55		Reaches new section
1.58		Reaches new section
2.03	No...I don't think so...I'm going to leave that question....	
2.08		Reads question 2
2.18	I've got a feeling I've just seen that when I was looking for .....fermentation	Scrolls directly back to top of document and reads the Contents
2.23	Basically the yeast dies....	Scrolls to relevant section and scans text.
2.37		Confirms answer and writes it down
3.04		Reads question 3
3.10	Something to do with Fermentation?....	Scans text, then scrolls down
3.19	Yes..to keep the yeast alive and stop the wine burning....	Reads text
3.21		Writes down answer
3.41		Reads question 4
3.47	Again, I think I've seen something on that	Scrolls continually down while scanning very quickly
3.55		Scrolls further down
4.00		Still scrolling and scanning rapidly
4.07		Still scrolling, has moved into previously unread text.
4.11	Ah...I must have passed it ...	
4.19		Scrolls back through the text scanning rapidly
4.24		Still scrolling back through the text

TIME	COMMENT	ACTION
4.28		Has scrolled back as far as Fermentation section
4.32		Has scrolled back to Introduction
4.37		Reading Contents
4.44	There's nothing in the Contents that's telling me.....but I'm sure it must be near the start of the document....It's fundamental.....	Reading the early part of the Introduction
4.53	Something to do with colour maybe....	Still reading Introduction and following sections. Scrolling down as necessary.
5.10	Bit I don't know what you mean by "lighter" (a ref to the question) or "taste"	Scrolls further don the introduction
5.23	..Ah.. it's the caramelisation of residual sugar	Quotes from the text having found a suitable answer
5.32		Writes down answer
5.51		Reads question 5
6.07	To the Index then....I haven't seen anything on this	Scrolls directly down to the bottom of the document
6.11		Scrolls slowly back up to the top of the Index
6.17		Scrolls quickly to the body of the Index
6.19	Grapes on page 1?....No..	Reading Index terms
6.25		Scrolls directly up to the top of the file and then scrolls slower down to a section in the introduction
6.56		Starts scrolling back through the Introduction
7.01	It must be in the body of the report then..	Reading section on Fermentation again.
7.03		Scrolls down to Aging section
7.06		Scrolling and reading the following sections
7.21		Studying the text intensely
7.27		Reading sections on Sweetness and Body. Scrolling slowly as required
7.31		Reading section on wine categories: table and dessert wines
7.33	Oh...dessert wines	Writes down answer
7.47		Reads question 6
7.51	I've just passed a section on aging	Scrolls up to Aging
7.55		Reads through section
8.09	Mentions a bit about vintage port...doesn't say how old it should be though.....	
8.22	I think I;ll find the section on Port	Goes straight up to Contents
8.25		Browsing through Contents
8.29	No...Index	Drags scroll bar down to end
8.32	Port....page 4	

TIME	COMMENT	ACTION.
8.38		Drags scroll bar to top
8.43		Selects Goto command from the menu. Inputs Goto Page 4.
8.45		Views sections on Table and Dessert wines
8.51		Scrolls don to section on Aperitifs and Fortified wines
8.53	Port...	Finds relevant reference
8.59	Vintage port...at least 20 years old	Writes down answer
9.14		Reads question 7
9.17	Haven't seen anything on this method before...Solera..Check the index	Drags scroll bar down to end
9.23	What a crappy index	
9.36		Drags scroll bar back up to middle of text.
9.42		Reads serially through the text from section on Aging to section on sparkling wines, using slow scroll as necessary.
10.30		Drags scroll bar to top to see Contents
10.34	I've just remembered..I can search for..	Invokes search facilities
10.36		Inputs "Solera"
10.52		Finds the appropriate answer
10.54	Oh..I missed that...I skimmed past it.	Writes down answer.
11.03		Reads question 8
11.07		Drags scroll bar to top to read Contents
11.09	My god...I'll search for that again.	Invokes search facilities
11.11		Inputs "Woodworm"
11.21	"Continue from beginning....?" yes	Hits return
11.25		End of document message. Search is unsuccessful
11.27		Reads question again
11.35	Wormwood...bloody hell	Corrects search term
11.55		Starts search
12.01	Vermouth eh.....	Term is found in relevant section
12.03		Writes down answer
12.10		Reads question 9
12.12	That's got something to do with champagne wines	There's a reference to sparkling wines as present position in text. Reads this.
12.24		Scrolls down text. Continues reading.
12.29	Produces natural effervescence	Writes down answer
12.51		Reads question 10.
12.54		Scrolls directly to top of text to see Contents.
12.56		Scrolls slowly through Contents

TIME	COMMENT	ACTION
13.03		Invokes Goto facility. Inputs "10"
13.07		Reading section on Rhone
13.20	Mainly red wines are there...	Reads through to Loire section
13.25.	Oh yeah..mainly white.a bit of luck	Writes down answer.
13.38		Reads question 11
13.44		Scrolls directly to top to view Contents
13.50	Countries start on page 6	
13.54		Invokes Goto command. Inputs "6"
14.01		Reading section on France
14.04	France, fairly obviously..	Reads through relevant sections on countries
14.24	France and the US, that is in the Italy section	Writes down answer.
14.39		Reads question 12
14.42	Types of wine?..	Scrolls slowly back through the text.
15.00		Scrolls slowly back reading text as he does so.
15.47		Reaches Contents
15.52	Ah...aging on page 2?	Scrolls back down
15.59		Reads section on aging
16.18	OK..it must be those.....	Makes notes
16.55	I think that must be the reference to it but it's not the same as "as soon as it's bottled"	Refers to question
17.10	Champagne...if you naturally ferment it, it's just going to stay in the bottle..	
17.20		Jumps down to Index
17.24		Scrolls slowly back through the Index.
17.30	What about Beaujolais	
17.42		Invokes Goto facility and inputs "11"
17.53		Reads sections on Burgundy and Beaujolais
18.20		Scrolls back on section just read
18.24	I'll look for bottled or something	Invokes Search facilities and inputs "bottle"
18.31	No..	Cancels this action
18.38		Reads sections on Loire and Burgundy
18.43		Invokes search facilities again and searches on "bottle"
18.48		Prompted to continue search from beginning of file he cancels.
18.52	No....	Tries search again using same term

TIME	COMMENT	ACTION
19.06	No..this isn't it	Has found numerous references to "bottle" in section on Aging.
19.13		Cancels Find command
19.19		Invokes search facilities again and searches on same term.
19.21		Reads the section on Port and Sherry where search facilities have taken him.
19.35		Find Next
19.37	Germany and Italy?	Find Next
19.41		In Champagne section. Find Next.
19.45		Is in California section. Find Next then takes him to Sparkling Wines section.
19.50		Find Next returns the "start from beginning message". He cancels the Find command.
19.54	So it's just champagne and semi-sweet, but that doesn't seem quite right.	Writes down answer. Session Ends.