Design and Delivery: Functional Colour Web Pages

Sahithi Pokala Siva

THE UNIVERSITY
of LIVERPOOL

Thesis submitted in accordance with the requirements
of the University of Liverpool for the degree of
Doctor of Philosophy

March 2001

Design and Delivery: Functional Colour Web Pages

Sahithi Pokala Siva

Abstract

It is nearly two decades since the early principles of designing graphical interfaces were formulated for the STAR computer systems. Today, the Web has revolutionised the way users interact, and consequently, the design of interfaces to match users' has changed. The interface arising from STAR's early principles seems to have led to continued redesign, and for all the psychological and other theoretical work in Human Computer Interaction (HCI), it seems that many potential display options have been ignored, or at least under-exploited. This thesis attempts to reconsider what is available and makes the suggestion that if HCI theory fails to generate appropriate uses for these options, then visual features such as background colour should be assigned a functionality arbitrarily. The success of this approach is dependent on such conventions becoming commonly adhered to. This thesis argues for colour as an example of an under exploited display feature mainly from the functionality standpoint. Other features that might also be considered under exploited include spatial location, apparent depth and motion.

Despite the growing research, the two main design problems that HCI has been facing are (1) knowledge and (2) the delivery means. This thesis suggests that the HCI research should be about delivering methods or design solutions that can be appropriately used by system developers who are not HCI experts. The thesis has made an attempt to deliver a colour product that could be used by both the Web page designers and end users on the Web. Firstly, colour being a popular visual feature, the potential Web Page Designs (WPD) are explored. Secondly, delivery requirements are taken into account in proposing the colour product, a three colour scheme, for use in WPD. These requirements include: learnability, simplicity and consistency.

The Web colour scheme has used colour to make it easy for users to make requests for the right kind of data and to promote easy browsing through the Web pages. The design of the scheme comprises a plausible classification of different types of Web resources into three categories, and a mapping of the simple perceptual category colour onto people's understanding of the semantically complex resource category. The resources are therefore colour coded to allow users to distinguish various types *at a glance* - that is by sight alone, without moving the cursor for textual explanation.

The thesis has employed a careful empirical approach in order to evaluate the effect of the proposed colour scheme on users' performance. Its robustness is validated with respect to its delivery requirements in a series of experiments.

Acknowledgements

I would like to thank Prof. Dan Diaper for his supervision and help in guiding this thesis through all its tribulations. I am very grateful to Dr. Trevor Bench-Capon for his encouragement, advice and support in completing the thesis. My sincere appreciation for Prof. Alan Gibbons for his timely help in continuing the research work.

I would like to thank Paul Leng and Prof. Lindsay MacDonald for their time and valuable suggestions for improving the work reported in this thesis.

My deepest appreciation for Kanti, Stuart MacGlashan, Paul Devine, Aideen for going through the draft, suggesting corrections and helping through all the stages of the thesis preparation.

Apart from this, the social ambience of the Computer Science Department made my stay memorable. I sincerely thank the staff and all my PhD colleagues for their assistance in technical and academic matters. The students, without them, the empirical work which I consider to be the main contribution of this thesis, could not have been undertaken.

I thank Prof. Jim Yip, Prof. Ian Haines and Howard Feather - the staff of the School of Multimedia and Informatics, University of North London for their support in completing the thesis.

I immensely thank my beloved family members for standing by me in all my endeavours.

Trademark Acknowledgements

Microsoft Internet Explorer, MS-DOS, Windows and Windows 95 are registered trademarks of Microsoft Corporation.
Java is a registered trademark of Sun Microsystems, Inc.
Netscape is a registered trademark of Netscape Communications Corporation.
HyperCard and MaCintosh are registered trademarks of Apple Computer Incorporated.
SuperCard is a registered trademark of Allegiant Technologies, Inc.
Xerox is a trademark of Xerox Corporation.
UNIX is a trademark of UNIX System Laboratories, Inc.
Hewlett Packard is a trademark of Hewlett-Packard Company.
IBM is a registered of International Business Machines Corp.
Inc. and Inc. Magazines are trademarks of the Goldhirsh Group, Inc.
Yahoo is a registered trademark of Yahoo Inc.
Alta Vista is a registered trademark and service mark of Digital Equipment Corporation.
Inforseek is a registered trademark of Infoseek Corporation.
NCSA is a registered service mark of Bales, gates & Associates, Inc.
MacWrite is a registered trademark of Apple Computer, Inc.
Flash is a registered trademark of Macromedia.
UIMX is a registered trademark of .
OSA is a registered trademark of Topsource Instruments, Inc.
OA is a registered servive mark of Eberbeck, Inc.

All other brand names and product names included in this book are trademarks, registered trademarks, service mark or tradenames of their respective holders. They are used throughout this thesis to identify the Web sites of their respective holders. No such use, or the use of any tradename or trademark, is intended to convey endorsement by the holder thereof of this thesis or any other affiliation with this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Prologue

1.1 Introduction

Recent developments in the design of human-computer interfaces are rapidly changing the way users interact with computer systems. Getting the interface right is becoming critical to the success of vendor products as users prefer simple but functional (functionality being a supplier driven factor) and easy-to-learn-use products at a small training cost [Bewley *et al.* 1983].

Under-Exploited Interface Features: At the heart of User Interface Design (UID) is the endlessly contextual and interactive nature of display features, for example, menus, icons, colour, layout and transient events (flashing, moving) etc. [Robertson 1979, Lansdale and Ormerod 1994, Travis 1991]. Considerable effort has been expended, both theoretically and in practice, on UI features such as menu design and icon appearance. Other User Interface (UI) features appear to have been under-exploited. Such features might include spatial location, motion, apparent depth, and colour [Diaper and Sahithi 1995]. In other words, UI and Human-Computer Interaction (HCI) research appears to have ignored potentially effective and perhaps cheap means of presenting information to people using visual features that already exist on most computer systems.

Among the features mentioned above, motion and depth need to be specially programmed to convey certain types of information. Colour, on the other hand, is readily available and does not require much specialised effort in programming its usage [Hunt 1994]. Further, psychologists have established that colour is a powerful visual cue (see Chapter 2). Despite this, little concrete research data is currently available on how colour features should be used in designing human-computer interfaces [Diaper and Sahithi 1995]. This being the case, the research in this dissertation concentrates on exploiting the potential use of a single feature "colour" in designing visual interfaces on modern computer platforms, in particular, on the Web. Very similar research could be undertaken with the other under exploited features and no claim to exhaustiveness is made of this features list. Appendix A.3 includes brief discussions on spatial location, motion and apparent depth.

Overall Objective of the Thesis: To investigate functional uses of colour in visual displays i.e. whether it is possible to use colour to communicate complex semantic information to computer users. The research is part of a larger programme of providing general, novel functionality to users by using already available, under-exploited, user interface features.

Research Aims: a) To establish that colour can be used to signal semantic information in a realistic human-computer environment; b) To test whether complex information can be conveyed to users through colour; c) To test whether semantic rules associated with colour are easy to learn and remember and whether such rules can be inferred; d) To evaluate specific designs that use colour in a semantically complex way and optimise their delivery; and e) propose designs (i.e. products of the research) for designers who are not HCI experts.

1.2 Designing Colour Interfaces

> "Colour displays are almost ubiquitous" [MacDonald and Vince 1994].
>
> "Colour is replacing monochrome in software packages, in display devices, and in desktop publishing facilities" (Hunt; taken from [Jackson *et al.* 1994]).

Over the last decade computer systems have grown considerably in sophistication and one of the main areas of change visible to users has been the continuing development of colour peripherals. CRT monitors have not changed much in terms of the basic technology but they have become easily affordable for a 'general' user, more reliable and available in a variety of convenient sizes for desktop use. What has changed radically is the cost and size of graphic cards to drive a monitor with high-quality 24-bit graphics. Simple palette-mapped 8-bit graphic cards are included as standard equipment on the motherboard of most personal computers; 24-bit graphic cards however are rapidly replacing 8-bit ones as a result of reduced costs. Consequently, hardware is no longer a limiting factor in the decision to choose colour on a visual display and users now perceive its absence as very old-fashioned.

Studies have reported a marked preference for colour displays [Schutz 1961, Christner and Ray 1961, Jeffrey and Beck 1972, Chase 1970] even though in some circumstances there was no measurable advantage in user performance [Munns 1968, Haubner and Benz 1985]. But what progress has been made on the principles that could drive interface design, both on traditional UI and on Web platforms, concerning the use of colours in a visual display for the purpose of supporting end users during task performance? As can be seen from looking at a current PC with Windows, the basic principles of the desktop metaphor have remained largely unchanged since the original Star system was introduced in 1981 [Smith *et al.* 1983].

The Web Platform and Colour:

> "The Web is like an amoeba on a very large scale - it continually absorbs its contents through hypertext links and changes its shape as a result of these new connection." - Jean B Gasen [Gasen 1995].

The popularity of the Web is rapidly shifting vendor competition from the traditional UI platforms to the Web platform. That is, from traditional GUI based software to the Web-based software enterprise, which is increasingly the environment of choice for the delivery of new products and for the linking of diverse systems in the academic, government and business sectors. The design of graphical interfaces on the Web has adopted the design principles and the basic components of WIMP (Windows, Icons, Menus, Pointing device) and WYSWYG (What You See is What You Get) user- computer interfaces. These interfaces, in turn, are based on the pioneering STAR system's [1] design principles.

The visual design of Web software adopted traditional UI principles but its development differs at many other levels, for example at the metaphorical level. The non-linearity associated with the Web demands new models, perspectives, tools and methodologies. Perhaps the most challenging task for the HCI community is to extend the existing models and tools or devise new ones and integrate them into the existing mosaic of the Web design.

[1] The Xerox STAR Introduced in 1981 is generally held to be the first GUI. STAR was articulated as a set of design principles by Smith *et al.* [Smith *et al.* 1983] based on psychological user modelling for the design of the WIMP and Direct Manipulation (DM) user-computer interfaces [Shneiderman 1983]. Since then there has been an exponential increase in computational power and work on HCI. Many of the successors of STAR are largely influenced by its design principles (Fig A.1 in Appendix A graphs STAR influenced systems [Johnson and Roberts 1989]). Today one area of the vendor competition is at the design of well-defined "look and feel" Graphical User Interfaces (GUIs) on both WWW and traditional UI platforms, which is likely continue to be the case in the near future [Tullis 1981].

As in the case of traditional UI platforms, the knowledge available on the the functional use of colour in designing Web pages could be questioned in the context of supporting users during Web browsing. There are several tools that allow aesthetic colour selection, but few that assist interface designers and common users in co-ordinating all the possible benefits of colours usage, and those that address the problem are very limited in scope. Colour usage is already exploited to some extent in Web design, most importantly to indicate hyperlinks and to distinguish those that the user has recently visited. The thesis investigates the further use of colour to convey semantically complex information to 'general' Web users prior to the selection of a hyperlink on a Web page.

1.3 Colour Interfaces - HCI Issues

Design principles, such as the 8010 Star Information System's, are the products [2] of early research in UID principles. These products were developed and delivered to the software industry so that they form valuable reference points and guide UI developers throughout the design process. As described by Smith *et al.*, these principles were neither simple in conception nor easy to apply to particular design issues. Delivery requirements, for example simplicity, therefore play a vital role in product design. The work presented in this thesis emphasises that design principles, in general, need to be expressed so as to be straightforward to apply in practice.

While it is not possible to cover the entire scope of HCI in this chapter, the following sections are intended to introduce two relevant issues (see Appendix A.1 for other issues) that have a bearing on the work presented in this thesis. As mentioned earlier, the Web-based applications such as the Web browsers and the design of Web pages are still rooted in WIMP and DM design principles, although the concept of a computer has shifted to the "Network is the computer" - as claimed by Sun Microsystems [CNN 2001].

[2]Following Long's terminology [Long and Dowell 1989a, Long 1986a], the thesis makes reference to discipline knowledge in terms of products of research which may assume different forms; for example, tacit, formal, experimental, codified - as in methods, principles, software tools, demonstrators and prototypes, etc., and even theories. It is not intended that products be physical entities, nor that they necessarily be manufactured or sold.

1.3.1 Design Principles

The Star Information System was heralded as starting a new age and style of computer use [Baecker and Buxton 1987]. While Star itself was not a long term commercial success, it was the foundation for the development, via Lisa, of the enormously successful, and now widely copied, Apple Macintosh interface. Smith *et al.* [Smith *et al.* 1982] claim that "Xerox devoted about thirty work-years to the design of the Star user interface" and that "the paramount concern was to define a conceptual model of how the user would relate to the system". Star's interface was based on eight principles which Smith *et al.* describe as their "main goals". These principles were:

1. Familiar user's conceptual model
2. Seeing and pointing versus remembering and typing
3. What you see is what you get
4. Universal commands
5. Consistency
6. Simplicity
7. Modeless interaction
8. User tailorability

Smith *et al.* make a plausible appeal to a variety of scientific research, particularly from cognitive psychology, in selecting and justifying their principles. The principles are not all specified at the same level of abstraction, they are not prioritised and they will often be in conflict. Figure A.1 does not show all successors of the STAR and is not complete (see section A.2 for the updated list). Many of these early design principles, however, have become implicitly enshrined in the design of human-computer interfaces by some mechanism of cultural diffusion, if not by simple copying.

HCI approaches have not given serious thought to delivering a set of general user interface design principles over the last two decades. Large "guidelines" or "style guides" [Hewitt and Gilbert 1993, Ravden *et al.* 1989] are self evidently not the children of a set of principles such as those proposed by Smith *et al.*, but rather have turned out to be mostly a collection of heuristics that designers believe have been useful to them in previous cases. Löwgren and Laurén [Löwgren and Laurén 1993], for example, describe them as "collections of advice".

Consideration of detailed research and evaluation work in HCI may lead one to the conclusion that people and their interactions with computer systems in different environments are very complicated and diverse so that it is neither possible nor desirable to develop general interface design principles, not even the ones that are expressed declaratively. Diaper [Diaper 1989b] argues that the primary product of HCI research should not be "declarative specifications" such as guidelines but "... *methods* that can be appropriately used by system developers who are not HCI experts."

1.3.2 Functionality

Smith *et al* [Smith *et al.* 1983] described functionality as the most important ingredient of the user interface and said:

> "... the best user interface work in the world may be wasted if it is not coupled with a product that offers significant functionality at a reasonable price, ... if there must be a trade off between functionality and the user interface, functionality must be given higher priority."

Hammer [Hammer 1984, Baecker and Buxton 1987] says:

> "The following list indicates the actual relative importance of various aspects of an Office Automation (OA) system:
>
> 1. Functionality
> 2. Functionality
> 3. Nothing
> 4. Functionality
> 5. Everything else
>
> ... Vendors have learned a lesson about ease of learning, ease of use, user friendly interfaces, and the like, but unfortunately it is the wrong lesson. ... They have forgotten two laws ...: first, people buy systems for functions,

> not for interfaces; and secondly, when a vendor talks about interface, it means he has nothing else to say.
>
> In reality, user interface is a second-order factor. All other things being equal, the system with better functionality will almost always win over one with a better interface ..."

Brown [Brown 1988] and Nielsen [Nielsen 1993] stress, however, that functionality is only effective if provided in a convenient and usable form.

Unfortunately, the concept of functionality appears to be a confused one. Diaper and Addison [Diaper and Addison 1992a] suggest that it is necessary to divide the concept as used in the literature into:

1. Human functional requirements and
2. Computer functional requirements.

Benyon [Benyon 1992], for example, clearly identifies functionality with the needs of people. Conversely, Diaper and Addison say:

> "However, there remains today a computer-centric view of functionality
>
> [Davis 1990, Collin 1989, Blethyn and Parker 1990, Pfleeger 1991].
>
> For example, Collin provided the following dictionary definition:
>
> *Functional specification: a specification which defines the results which a program is expected to produce."*

Diaper [Diaper 1989, Diaper 1990, Diaper and Addison 1992] has previously suggested that the major problem facing HCI today is how to bridge the gulf between anthropocentric user requirements and computer-centric design. While his own approach has involved specifying user requirements using formal notations [Diaper and Addison 1992a], an alternative, more procedural approach has been described by Carver [Carver 1988]. She used JSD (Jackson Structured Design) to design "screen layouts" with the intention of starting computer system design with structured interface designs which would then drive the design of the back-end computer system [Cameron 1983]. It is unclear why this general approach, ignoring the JSD part, has not been more fully explored in recent years, particularly given the enormous increases in computational power that have occurred since her work.

Considerable effort has been expended on icon appearance. In this context, the Web and UI researchers have been concerned [Nielsen 1993, Nielsen 2000] with the need for evaluating icon design with respect to functionality. It should be self evident, however, that if the functionality triggered by an icon is not what users want then no amount of work on the design of the icon's appearance can correct such a functionality fault.

It, of course, is possible to separate functional interface design from the implementation details of the interface. One obvious area where functional interface design has not been supported by the visual appearance is with respect to logical menu design. Apperley and Spence's [Apperley and Spence 1989, Anderson and Apperley 1990, Cockton 1990] "Lean Cuisine" notation provides a functional menu specification language. Using the "MacWrite" style menu they illustrate how the logical description of a menu fails to represent the user's real options concerning menemes (menu options) with respect to: selection and deselection; default options; and exclusivity or compatibility. It is still the case that some of these functional properties of menus are not fully visually represented to users. This indicates a need for an internationally adopted stylistic convention or standard, similar to the accepted style of showing non-selectable options by greying them, or to how cursor appearance changes to indicate processor state changes (e.g. while the processor is busy, various window managers indicate this to users by changing the cursor to: an hour glass, an analogue clock, or a busy bee).

It should be of serious concern to those who design the visual appearance of interfaces that the current conventions do not support even the currently available functionality. That this type of visual design error is still with us can be taken as evidence of support for Carroll's criticisms [Carroll *et al.* 1988] concerning the design by redesign approach inherent in participative, rapid prototyping.

1.3.3 Consistency

> "The choice of colours for a GUI is less important than their consistent use throughout all screens. You should make every effort to achieve both internal consistency within the application and external consistency with other applications, systems, and real-world conventions" [MacDonald 1999].

From Neilsen [Nielsen 1989] and Shneiderman [Shneiderman 1992] to MacDonald [MacDonald 1999], researchers and practitioners have published articles to demonstrate the role, advantages and disadvantages, and complexities associated with implementing different types of consistency in system design [Tero and Briggs 1994, Rosenberg 1988, Lee and Fisk 1993, Lee *et al.* 1994, Sankar *et al.* 1995, Grudin 1992, Neale 1993, Gupta 1986, Grudin 1989, Grudin 1990, Tanaka *et al.* 1991]. For example, over the last decade "Strive for Consistency" has been constantly listed by Shneiderman [Shneiderman 1987, Shneiderman 1992, Shneiderman 1997] as the first of his eight golden rules for designing user-computer communication. Guidelines and style guides for user interface consistency have been discussed in human factors and software development areas and the cost benefits associated with achieving consistency have been widely acknowledged [Simpson 1999]. Over the past five years the need for achieving consistency in the dynamically changing world of the Web has been addressed [Lynch and Horton 1997, Newman and Lamming 1995].

Consistency being a key aspect of usability in user interface design [Shneiderman 1992, Shneiderman 1994], many conferences and workshops have been held but have failed to give a precise definition of the term 'consistency' [Neale 1993, Grudin 1989]. One prominent interface designer has acknowledged the difficulty associated with defining the term as "I know consistency when I see it" [Grudin 1989]. Some of the connotations that people assigned to the term consistency are:

- "Consistency is the similarity of related parts in such components as the user interface and the documentation" [Perlman 1989].
- "A consistent interface is Predictable (users anticipate what the system will do), Dependable (the system fulfils the user expectations), Habit-forming (the system encourages development of behaviour patterns), Transferable (habits developed in one context apply in new situations) and Natural (the interface is consistent with users' understanding). These attributes apply within an application, across applications, and even through complete product-lines" [Blake 1986, Koritzinsky 1989, Tero and Brigg 1994].
- "Consistency means that similar functions in an interface, such as choosing a single item from a list of mutually exclusive options, are represented and interact with the user similarly wherever they are found in an interface. Consistency does not mean that such a function is identical throughout the interface, just that where appropriate the technique is similar and where differences exist, those differences have meaning" [Wiecha 1992].

- "Consistency means that similar user actions lead to similar results. Consistency allows the user to transfer skills from one area to another" [Wolf 1989].

Owing to differences in user groups, problems, tasks and environments several facets of consistency have been defined at different levels - ranging from individual applications to products of the entire computer industry - in terms of different components, for example, visual appearance and behaviour (i.e. consistent look and feel). A consistent look can be achieved, for example, by having each folder icon on the desktop identical in shape and size and thereby establishing a pattern [Verplank 1986]. A consistent behaviour can be promoted by designing interface objects exhibiting predictable behaviour [Smith *et al.* 1982]. For example, when an object, such as a menu option, is selected it always reverses video. The same behaviour should be followed for all types of objects, text or icon, to establish a behaviour pattern [Koritzinsky 1989]. Existing literature describes (see Glossary) other components (conceptual, communication, physical) or classes (abstract - metaphorical; conceptual - role and categorical; stylistic - presentational and interactional) and different types (semantic, syntactic, physical, internal and external) of consistency [Kellogg 1987, Braudes 1991, Pangalos 1992] at different levels.

At each level consistency needs to be ensured, for example, in the context of several hundreds of appearance elements (e.g. colour, typography etc.) and associated interaction rules (e.g. rationale for pull-down or pop-up menus) to build consistent look and feel for even the simplist interface design. An evaluation of UI along all the facets of consistency would be useful to appreciate the trade-offs and priorities amongst varieties of consistency represented by a design.

Different methods (e.g. user-defined and centralised approaches) for achieving consistency have been defined, each promoting consistency to a certain extent. Having identified the role of consistency as a major design issue and after weighing advantages of consistency over its disadvantages [Nielsen 1989], major corporations such as Apple, IBM, DEC and AT & T have announced plans for standardising user interfaces both within and across product lines and therefore establishing a corporate standard. As part of this process several guidelines and corporate style guides are proposed to ensure consistency in order to increase the effectiveness of design activities. The inherent complexities associated with developing a set of standards (international and / or corporate) to build a completely consistent interface, however, have several implications. For example, financial implications are complex because large initial investments must be made in return for potential, long-term profits [Rosenberg 1989, Wiecha 1992].

In the past, a number of reasons have led to the belief that consistency is hard to achieve. Besides the positive efforts being expended to develop solutions to consistency problems, there are strong reasons against building consistency into user interfaces. Against Shneiderman's first golden rule [Shneiderman 1992], Grudin illustrates the importance of trading off consistency for other project objectives [Simpson 1999] through a set of well-defined cases such as menu default patterns [Grudin 1990, Grudin 1989]. These cases provide a necessary caution for designers to not to enforce "foolish consistency" in system / interface design and emphasise that consistency should never be applied as the foremost rule in every design situation. It should be given high priority and efforts need to be expended in building consistent user interfaces, and linking them to different user groups and tasks, as well as developing a better understanding of how consistency increases productivity in spite of its disadvantages in a few situations.

Nielsen [Nielsen 1989] says:

> "Consistency in user interfaces follows the second law of thermodynamics: If nothing is done, then entropy will increase in the form of more and more inconsistency in user interfaces".

Entropy is the measure of disorder of a system [Sands 1963]. If we accept this, it follows that mechanisms are required to meet user needs and challenges of rapidly changing technology for maintaining consistency across corporate-wide suites of applications to enable integration and allow users to perform effectively with a wide range of software products. This research identifies consistency as a primary requirement for delivering any software product to the users and investigates some plausible solutions for designing consistency for human-computer interfaces for both the Web (see Chapter 4) and traditional UI platforms (see Chapter 3).

1.4 Focus of the Research

The research reported in this dissertation has made an attempt to investigate current knowledge on the uses of colour. One only has to look at the index of most HCI text books to see how little is known about what colour should be used for, other than for simple grouping and to indicate an emergency. Brown and Cunningham's [Brown and Cunningham 1989] book is one of the exceptions in that it devotes a whole chapter to colour (26 pages out of 271, i.e. 9.7%). More typical are Helander's [Kieras 1988] HCI handbook, (less than 4 pages out of 1,167, i.e. 0.3%), and Galitz's [Galitz 1985](8 pages out of 218, 3.7%). Travis book [Travis 1991] on colour displays devotes only one chapter to coding and formatting aspects of colour (21 pages out of 190, i.e. 11.1%). Much of the material presented in this chapter duplicates the topics described by Brown and Cunningham [Brown and Cunningham 1989]. It is also the case that much of the material in most HCI textbooks that include anything more than a few sentences about the use of colour, concentrate on colour perception and the psychophysics rather than the potential use of colour [Corte 1986].

An extensive literature review on colour (see Chapter 2) has indicated that current research in this area amounts to little more than conjecture, rules of thumb and subjective guidelines for designing colour displays. Further, there exist cases where some of the guidelines contradict each other. For example, Nielsen [Nielsen 1993] recommends "try to use colour only to categorise, differentiate, and highlight, not to give information, especially qualitative information". In contradiction, Tufte [Tufte 1990] discusses using "colour as quantity" i.e. for measurement purposes [3].

Critical to the main thrust of this research is that it is possible that in UID, colour has not been used functionally because it is not a very good feature for communicating anything but simple concepts and perhaps, implicitly, researchers and interface designers have realised this (Nielsen's guideline). Certainly in the real world, colour functionality is generally used to categorially and the concepts signalled are simple (stop/go, hot/cold, etc.). Exceptions include use of colour in cartography where colour is used to signal concepts on a continuous scale. For example, colour indicates proportional change (e.g. altitude) but does not provide an absolute value. When colour is used continuously (e.g. to indicate temperature, say on a shower) then this is usually used redundantly with some other indicator such as spatial location. Surely, it would be difficult to imagine a situation where a designer replaces a car's pointer speedometer with a graduated colour one. The thesis intends to examine whether colour, even when used categorically, can inform users about complex system states; where "complex" means that more than one state property are combined (e.g. likely ease of access and likely usefulness of a Web resource, see Chapter 4).

[3] Nielsen's guidelines are mainly focused on the use of colour on hardware-variant visual displays. On the other hand, Tufte's guidelines on colour cover different media, in particular, print media and some of his recommendations are based on use of colour in Cartography where colour has been traditionally used to indicate features such as height above sea level, proportional density, etc.

In summary, the argument presented is that there is no shortage of human-computer interface features but that the R & D effort on exploiting the uses of the available, perhaps cheap, display features, has been unequal. The thesis focuses on colour for the reasons stated in section 1.1.

1.4.1 Functionality of Colour

Despite hundreds of person years of expensive effort having been devoted to the development of high resolution, large palette colour displays, little progress has been made on the principles that could drive interface design and the use of colour to give a meaningful visual display. Given the vital importance of functionality [Hammer 1984, Diaper and Addison 1992a] this point may be disquieting to the HCI community who have failed to provide an adequate theoretical base that could be used to address the question: 'What should colour be used for?'. The thesis addresses this issue with respect to the design of pages on the Web; what colour might signal and whether users can utilise consistent, complex semantic associations bound to colours. The main theme of the research therefore is to exploit the functionality of colour in designing Web pages [4].

In addressing these questions, the research will consider:

- The kind of use people such as Web page designers make of colour in their work.
- The way in which they go about colour design.
- The resulting impact on users.
- The delivery means of such work, i.e. instructing designers and end users on how to deploy and interpret colours.

1.4.1.1 HCI Research - The Web

The Web is evolving from an information repository into a distributed interface to a globally networked, computational engine [Staffordshire University 1998]. Millions of users are connecting their machines to the Web to exchange information, access remote resources and work together by sharing knowledge. In spite of these new trends in

[4] Initially, the colour usage issue is addressed with respect to the design of traditional UI applications as part of the initial research (see Chapter 3).

online information access and exchange, much of the theoretical work in HCI appears to be concentrated on a one- to-one relationship between the user and his/her computing environment [Almeida *et al.* 1995].

The Web poses several unique design problems to the HCI community, specifically in the context of functionality and usability. Users are exposed to a number of interaction problems that seldom occur with traditional computer system platforms. For example, the accessibility of information - a Web page generally contains a of number of hyperlinks pointing to a variety of information available on the Web either at a local site or a remote location. People are often faced with unpredictable timing delays when they attempt to access information in particular over remote networks [Johnson 1995]. Further, in the maze of different forms of hyperlinks, it is difficult for them to know the content of a page that they have selected for downloading. The desirability of content varies across Web documents. Links pointing to "under construction" pages, dead pages or pages with statistical data dumps present users with yet another class of problem.

Many Web users only possess a minimal knowledge of the communications mechanisms that support computer networks. Remote site failures, network delays and the problems of the representation of information have little relevance to their every day tasks. As a result, the problems that they face on the Web lead to frustration and error which can, in turn, prevent them successfully retrieving necessary information. It should be noted that many users are infrequent users of particular pages with no learning or training and need to use familiar conventions to complete their tasks. The functionality with respect to accessing world wide distributed information is crucial in WPD [Brown 1996]. It is therefore necessary for the HCI community to look into these design problems and propose solutions to match the technology to the needs of users.

In this direction, this thesis attempts to demonstrate some appropriate styles of solutions to some of the above mentioned problems. The solutions are in the form of designs that can be implemented by users (both Web page developers and end users) who are not HCI experts. The designs involved (1) thinking of how the information can be categorised and represented to users so that they can identify different types of information, *at a glance*[5],

[5]There may be many possible definitions of visual psychological moment, ranging according to definition from 20ms (Bloch's law) to seconds (subjective reporting). The definition of *at a glance* used in this thesis reflects the empirical result that stimuli such as colour, can be processed by the visual system in parallel if discrimination is based on a disconjunction of stimulus properties. The thesis makes a reference to Treisman and Gelade [Treisman 1982, Treisman and Gelade 1980]. They claim that grouping on colour occurs only in parallel (i.e. *at a glance*) when colour is disjoint from any other visual feature. In other words, users are not able to distinguish, *at a glance*, objects defined by one colour and one shape amongst

without actually traversing the hyperlink and also without even moving the cursor, and (2) proposing novel ways of using different features of colour, for example background or border, as important visual cues to represent the categories of information.

On the Web, colour coding hyperlinks is not entirely a novel idea. At present, the browser software such as Netscape and Internet Explorer uses colour to indicate three different states of links *at a glance*: visited (links that are already viewed by a user), unvisited (links to Web locations not yet viewed by a user) and active (links currently active or being visited). That is, a simple perceptual category 'colour' is mapped on to a simple function 'the state of a link'. It should be noted that this colour coding only applies to indicating different states of text links. Web browsers can only indicate the active state of a graphic link by displaying a message in the status bar. Currently, there exist no practices to indicate graphic links that are visited and not visited. Graphic links e.g. image maps also pose other navigation problems (see Chapter 4 for details), for example existing practices do not clearly differentiate selectable graphics from non-selectable ones *at a glance*. There are alternative practices, for example changing the shape of the cursor on rolling over a selectable graphic and cursor-over tool tips: when a user moves the cursor over a graphic the browser can display an appropriate message in a pop-up text box or trigger some appropriate action. Clearly, these practices do not provide information to users without moving the cursor around regions in a Web page. In this context, the thesis investigates whether background and/or border features of colour can be used to communicate information that is not already represented to users, *at a glance*.

1.4.2 Design and Delivery

The thesis has previously mentioned some HCI issues (see section 1.3 and Appendix A.1) relevant to both traditional UI and Web platforms. In particular, the discussion hinted at HCI research and developing methods for users who are not HCI experts [Diaper 1989b].

If the academic and industrial HCI research and development community are "to give HCI away" [Diaper 1989b] then there must be something for them to give and some means of giving it. This requires both a product and a delivery mechanism.

different, same coloured shapes, although they are able to perceive all objects of the same colour by ignoring shape). It is not known how this definition of *at a glance* might relate to other visual processes, for example eye movements. The phenomenon however is large, robust and particularly suitable for the realistic human-computer environments that concern the final experimental stages of the research work.

The work presented in this dissertation emphasises that the products, for example colour user interface principles, guidelines and demonstrations, have been designed so as to optimise their delivery. While it may be specious to suggest that the delivery aspects precede and drive the design of the products, not least because design itself is not a tidy process, at a minimum, the colour user interfaces developed have had their requirements heavily constrained by many aspects relating to their delivery.

Activities associated with product design and delivery can be described in a framework based on Long's [Long 1986a] framework of HCI models. Reasons for using Long's framework include (a) It is sophisticated and exemplified within HCI; and (b) it is well established [Long and Dowell 1989a, Long 1986a] and is still current [Cummaford and Long 1998]. Figure 1.1 represents a general framework of HCI activities associated with design and delivery of products to users in the real world. In the framework, the real world is contrasted with a representational world whose function is to facilitate the understanding and change of the real world. The two worlds are related by means of intermediary representations and associated activities, which transform one representation into another. In the figure, representations are shown as boxes and transformational activities as directional arrows. The Science Representation consists of the human sciences (principally both experimental and social psychology), computer sciences and other sciences that might include physical, biological sciences etc. HCI is concerned with people and computers. Scientific knowledge is applied to real world of people and computers by several transformations. Particularising the scientific representation produces a Research Engineering (RE) representation consisting of HCI (principles, practices etc.), UID, WWW and Software Engineering (SE; design principles, guidelines etc.), and others. Application of research engineering knowledge produces research engineering product representation, consisting of a product specification, design, implementation and evaluation.

The research presented in this thesis extends Long's framework and identifies three types of user population: research engineers, engineers and end users. Research engineers explore the knowledge to define the problem, prepare a specification, design and evaluation leading to the invention and develop/manufacture of tools to solve the problem. The tools could be theories, methods, design principles, guidelines etc. The outcome of research engineering (i.e. tools) is represented as Research Engineering Products (REPs) in the framework. REPs are delivered to engineers who use them or instantiate them in designing products (for example, software applications) for the end users i.e. Engineering Products (EPs). Producing an EP from a REP involves bringing together the user needs and other contextual factors with the REPs products supplied by the research engineer. Both REPs and EPs are studied in order to allow system, user and environmental constraints for a computerised system to be derived from the real world and then are integrated into an understanding of the phenomena of the real world.

Understandably for a new discipline [Sassie 1997] HCI has generally concentrated on the design and development of research products (REPs) which may eventually be used by the professional software industry (in the development of EPs), on the basis that the product (i.e. the first item in the list) is a necessary pre- condition for the delivery (i.e. the second item).

Figure 1.2 shows a variation of Figure 1.1 and illustrates a framework of activities specific to the current research. The thesis framework varies with respect to the types of user population and the delivery mechanism, most notably that the REPs were not delivered to designers, although this delivery was considered during their design. The reason for not carrying out this is that prior to delivering REPs to designers, it is necessary to test that the design examples derived from the REPs do actually work with real users. At this stage of development, the research engineer also performs the role of an engineer and develops both REPs (e.g. proposed colour designs, see Chapters 3 and 4) and EPs (e.g. Xaero in Experiment 2 in Chapter 7) and delivers them to the real world users. The REPs were instantiated in specific instances as part of applications already available on computer platforms (e.g. in a Web page and UI applications such as Xaero) and were tested for their influence on user performance through a series of experiments.

In particular, Figure 1.2 describes the work presented in this dissertation as a general framework of the relationships between its three major, conceptual components: Delivery, Products[6] and Experiments. This framework has been used in this thesis to explain the various experiments' rationales and to interpret their results. The design of the experimental materials, procedures and analysis have all had to meet the requirements directly generated from considerations about how to deliver the results of the research to the software industry and the end users. Further, the experiments must meet all the delivery requirements of the products that are being designed. The experiments provide a rich, evaluative function of the proposed products (see Chapters 3 and 4 for the UI colour scheme and the Web colour scheme, respectively) and the intended means of delivery. The introductory chapters of this dissertation review these and "other requirements" that are relevant from the HCI, colour and the Web literature. The framework also suggest that delivery issues directly effect the design of the products; again, as requirement constraints.

[6]The thesis refers to REPs as products unless otherwise stated. REPs in this thesis refer to UI and the Web colour schemes described in Chapters 3 and 4.

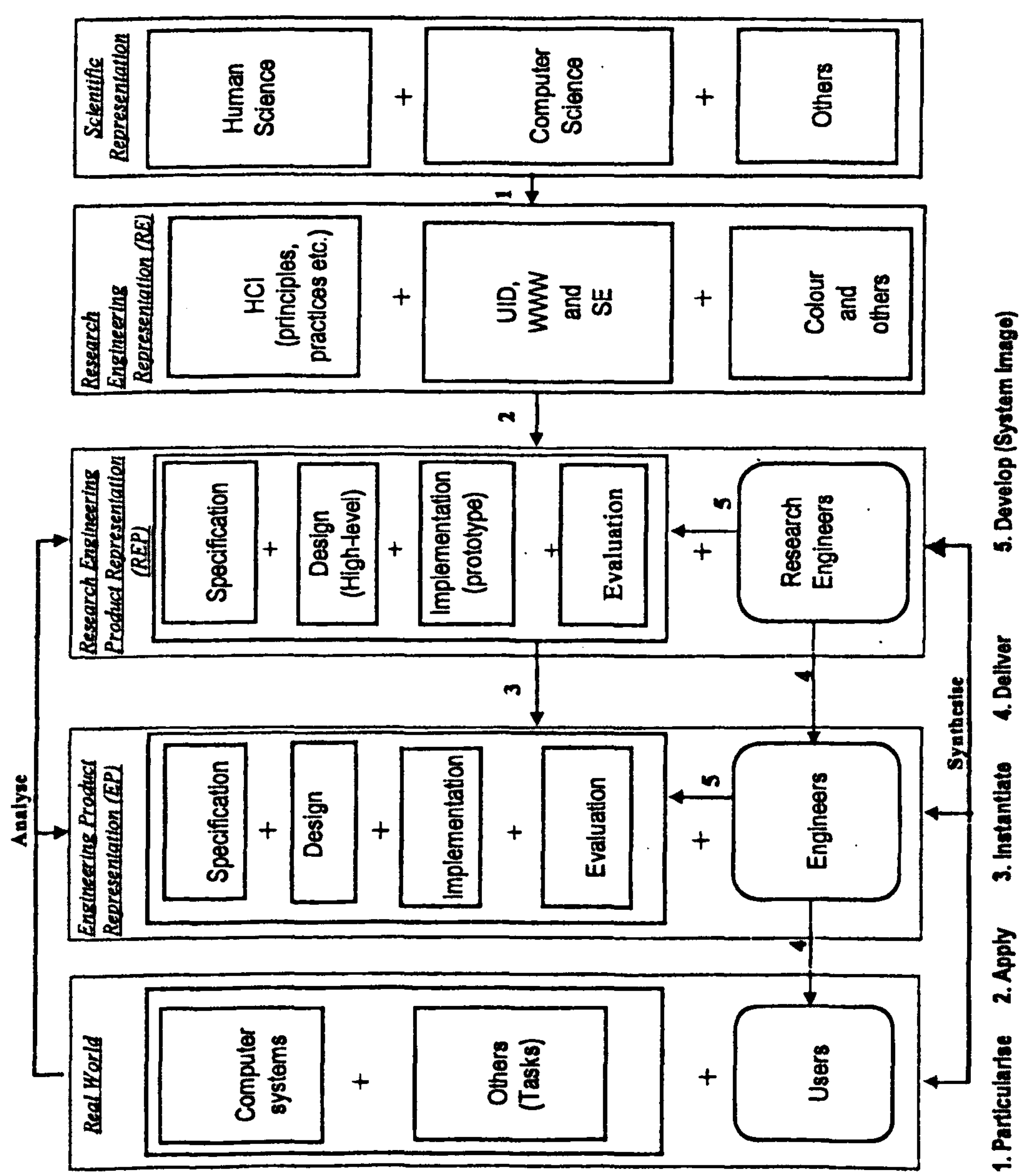
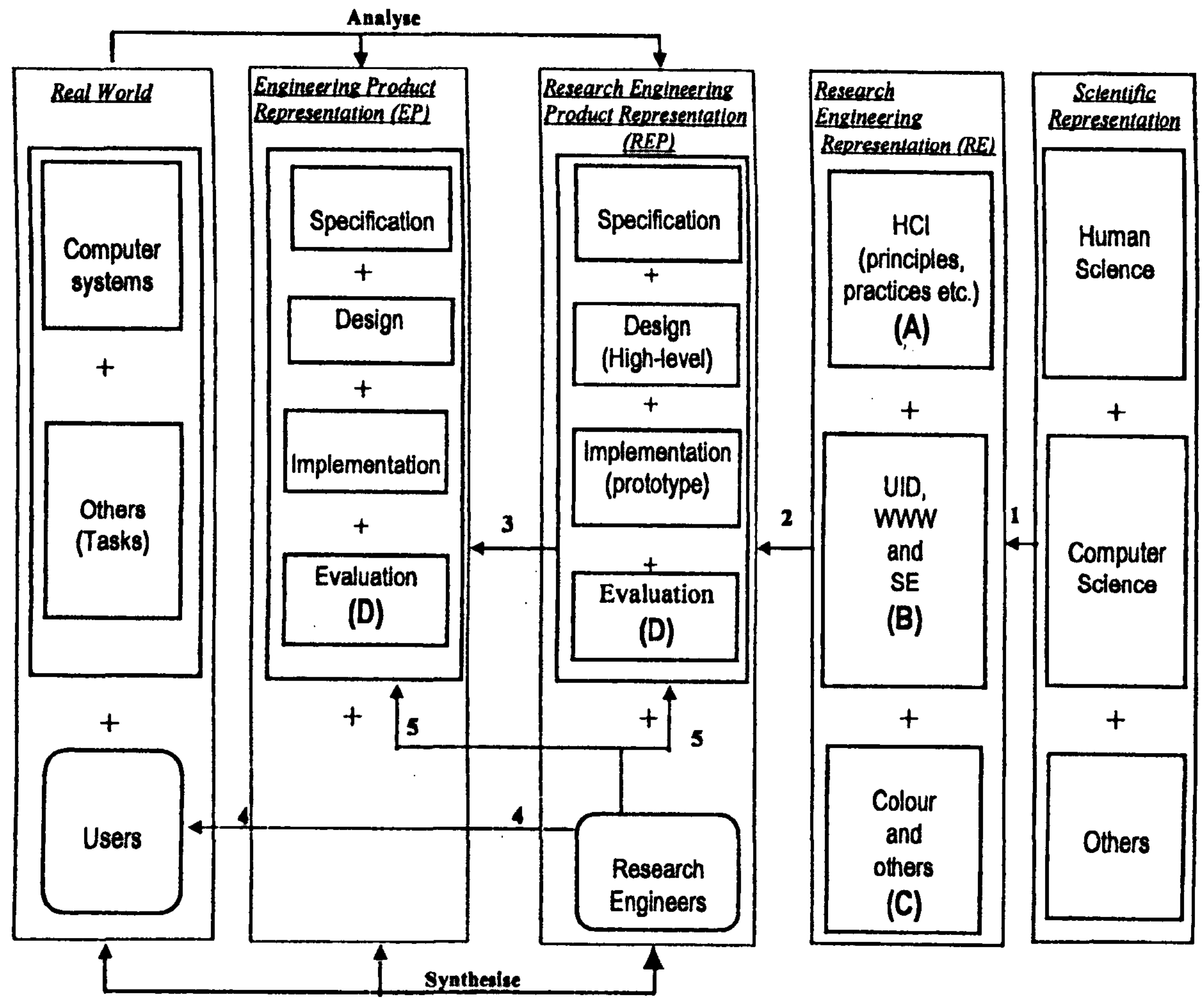


Figure 1.1: A general framework of activities associated with product design and delivery



1. Particularise 2. Apply 3. Instantiate 4. Deliver 5. Develop (System Image)

(A) Chapter 1 (B) Chapters 3 & 4 (C) Chapter 2 (D) Experiments 1 to 7

Figure 1.2: A framework of activities in this research

To illustrate the importance placed on the effects of the delivery aspects on the experiments and products, the primary product of the thesis (i.e. the Web colour scheme) has been restricted to three colours so as to make it easy to learn, remember and apply - these are example delivery requirements [7]. Notwithstanding comments from some users that it might be better if more colours are used, evidence is presented that supports the simpler design through a series of usability studies conducted to evaluate the use of colour (Chapters 5 to 12).

In addition to the Web colour scheme, the thesis also demonstrates another colour scheme designed for use on traditional UIs (see Chapter 3). Prior to delivering the Web colour scheme, the thesis examines users' ability to learn, remember and apply colour rules on a UI test platform. The UI colour scheme and corresponding experiments form the test bed for the Web colour scheme work.

Both the Web and UI colour schemes were designed to be independent of software and hardware platforms in order to make them usable by virtually everyone within the wide range of real work environments in which computers are used. In addition, the benefits of a functional use of colour will need to be quite large to have any chance of being adopted as a convention. For example, the use of red to indicate important or warning information is, virtually, a universal convention. It is nearly impossible to propose universal laws governing the use of colour in visual displays, but once conventions are established, they become de facto standards. Similar to red, the thesis proposes other colour meanings to be adopted in similar fashion.

1.5 Summary

The research reported in this thesis proposes a high-level design i.e. the Web three colour scheme (see Chapter 4) which suggests novel, functional uses for colour on the Web, in order to provide information that was not already available on most Web pages. The creative part of the design (not just an iteration of previous designs) involved thinking of what additional information users might find helpful during Web browsing. The thesis also presents another colour scheme, the pilot product of the thesis, for use on traditional UIs. Both the Web colour scheme (i.e. the main product of the thesis) and the UI colour scheme (see Chapter 3) exploit the proposed functionality of colour when used consistently both

[7] The restriction on the number of colours to three limits the number of associated categories described in Chapter 4

within and across Web pages and traditional UI platforms. The story of the usability and effectiveness of these products is told in the series of usability studies. The purpose of the experimental work described is to verify hypothesis about the effectiveness of these products. The experiments are presented in chronological order. The initial experiments evaluated the sensitivity of test measures on user task performance and those that followed tested the effect of the designs on users in a variety of working conditions.

While one could undoubtedly train people to use colour functionally (i.e. where colour is meaningful in that on a user interface it signals some state of the system), the aim of this research is to investigate whether people with little information or training can make use of it. As mentioned in the previous section, to cater for nearly all computer users it is necessary that the use of colour is platform independent and robust across the range of environments in which computers are used. This has implications for the designed use of colour. In order to be platform independent, it is necessary that relatively few, distinct colours are used. In the experiments, apart from the colour being used categorically, i.e. red, yellow, green and blue, the population tested are - within the available resources - heterogeneous and the environmental conditions should not be important.

1.6 Organisation of the Thesis

In Figure 1.2, the work presented in this thesis is clearly mapped on to the associated components/ activities in the framework, both graphically and in text, in order to identify the contributions of the different chapters concerning literature review and those describing experiments.

Chapter 2 is intended to provide the reader with an overview of the different aspects of colour in the design of computer systems relevant to the thesis. The chapter describes the application of colour in current interfaces and reviews previous research on principles, standards, guidelines and styleguides, concerning the functional use of colour.

Following Chapter 2, the emphasis shifts to focus on general functionality and consistent usage issues. Chapter 3 provides background on consistency [Shneiderman 1992, Nielsen 1989, Smith *et al.* 1983], one of the delivery requirements that drives the design of the colour schemes reported in this thesis. Further, the chapter provides the details of the pilot UI colour scheme and illustrates the role of colour in promoting consistency.

Chapter 4 introduces the main product of this research i.e. the Web three colour scheme.

Different issues related to the Web are discussed in order to provide sufficient background for the delivery and design of the Web colour scheme. The functional aspects of the scheme are discussed with respect to some common problems heard from the Web users. Chapter 5 provides a brief introduction to the experiments presented in the thesis and describes related issues. The chapter discusses use of discount usability methods, for example simplified think-aloud [Nielsen 1993], as part of the empirical work and the reasons behind conducting the first three experiments on the traditional UI platform and the last four on the Web platform.

Chapters 6 and 7 present the details of Experiments 1 and 2, respectively. They were conducted to evaluate the pilot UI colour scheme and include examining whether people are sensitive to the colour rules and whether the proposed functionality of colour has an effect on their task performance. Prior to testing for differences between experimental conditions, it is desirable to demonstrate that the experimental design is able to detect differences between conditions. Experiment 1 was therefore aimed at testing the sensitivity of some of the test measures of users' task performance. The UI colour scheme was applied consistently both within and across application screens, and evaluated under colour and monochrome conditions. The results indicated that test measures, for example task completion times, are capable of detecting differences between test conditions. Test users in consistent colour condition took less time in completing the test task than those in monochrome. It is probable that users did learn colour rules and there is positive transfer of colour training. The experiment however does not prove what is actually learned because the experiment was not specially designed to test independently positive and negative transfer of colour training effects.

Experiment 2 examined the colour scheme for positive effects on users' performance in realistically complex tasks under different colour conditions: consistent, inconsistent and monochrome. It used more test conditions and users compared to Experiment 1. The study failed to show any difference between test conditions, which means that the task performance was independent of colour or even monochrome. The study however demonstrated some important results in that while the colour UID scheme is plausible, it cannot, at least some times, be usefully applied to real application interfaces. Following Experiment 2, testing of UI colour scheme was abandoned for several reasons (see Chapter 7 for details), the main reason being the lack of generality of its application.

Chapter 8 describes the third experiment, which was designed based on the failure of the second experiment. The experiment was aimed at testing lab based results from Treisman and Gelade [Treisman and Gelade 1980] on use of colour for grouping screen elements in a computer realistic setting. The experiment demonstrated the predicted effect of colour, that is, colour allows users to selectively attend to the target stimuli and ignore the non-target ones by reducing the psychological size of the search space. The experiment therefore was a success in testing the crude grouping effect of colour and the sensitivity of people to colour organisation. Further, the study established users' ability to learn and apply a single colour rule.

Chapters 9 to 12 describe four experiments conducted to validate the delivery of the Web colour scheme. All the experiments were designed to test the colour scheme when applied consistently both within and across Web pages. Chapter 9 details Experiment 4, the purpose of which was to extend design presentation (buttons) used in Experiment 3 to background and border colours in Web pages in order to test the grouping advantages of colour. Further, the experiment was intended to test the difference between experimental conditions and develop a research platform for the new functional use of colour for designing Web pages. The study was a success and demonstrated that the colour grouping advantages are independent of their format.

Chapter 10 reports the fifth experiment as an extension of the previous one. The study tested the users' ability to learn and apply the three colour rules involving mappings between the simple perceptual category 'colour' and a complex function indicating 'the type of a hyperlink' (in terms of accessibility and content), and design formats (buttons and image maps varying in size and shape). The experiment demonstrated that after training, people could correctly apply at least three, complex colour rules that are graphically independent (backgrounds and borders). One important point to be noticed in all the studies reported here is the type of users. The delivery of the colour scheme (REP) was tested primarily with end users, but not with professional designers although this delivery was considered in its design (see Figure 1.2). As mentioned in section 1.4.2, before giving the products (REPs) away, the experiments attempt to verify and validate some design examples, generated by instantiating the colour scheme (REP) in several Web pages (EPs), do really work with end users who are generally not HCI/Web professional designers.

Chapter 11 explains the sixth experiment the purpose of which was to: a) demonstrate that users who have been taught a colour rule can apply it to different graphical formats; and b) investigate whether exposure can lead to implicit learning. This experiment was similar to the fifth one in overall design, but differed in the test task design. The task included specially designed graphics in order to encourage users to infer rules during their task performance. In the realistic working environment not all graphic images that are commonly found in the Web pages are designed by professionals. Since every end user is potentially a Web page author, the graphics used in this experiment (and other experiments) were purposefully not designed by professionals in an attempt to simulate a realistic Web browsing environment. The graphics also provided a strong test of explicit learning in that some antagonistic features (see section 11.2.1) were used which it is shown, can over-ride users' learning and application of colour rules. The experiment demonstrated that people cannot infer colour rules and therefore need to be told explicitly the meaning of each colour. An important result is that the presence of antagonistic features can affect users' ability to remember and apply colour rules. This means that colour in the Web design is plausible but that users' model of the colour scheme in the presence of antagonistic features, may be cognitively weak.

Chapter 12 describes Experiment 7, which includes design / redesign of real Web pages using the three colour scheme. The main purpose of this experiment was to measure subjective satisfaction. Users' were introduced to a colour style manual and were asked to complete a questionnaire. User's comments and the responses were summarised. Overall, the analysis of users' ratings indicated that the colour design had a positive impact on their assessment of its usefulness in browsing Web pages. Chapter 13 concludes the research and discusses the scope for further work.

Chapter 2

Colour Usage in Human-Computer Interfaces

2.1 Introduction

Colour and its associations form an integral part of people's lives. It can be a guide to the quality of food, establish a mood, represent realism, an effective danger sign, indication to individual personality etc. [Foley *et al.* 1990, Jackson *et al.* 1994]. Owing to its importance as a powerful visual feature, colour has been the subject of study for researchers and designers from various disciplines including cognitive psychology, HCI and UID.

Computer hardware developments provide neat bookends to advances in display technology that have made the use of multicolour display feasible for a variety of applications across several platforms. One bookend is the introduction of high-end performance workstations such as the Silicon Graphics (SG) workstation in the mid-1980s that brought advanced colour graphic capabilities and new visual effects at an affordable price. The other bookend is the absorption of these advanced colour graphics capabilities into the household PC platform. Despite the fact that current display technology is limited in terms of the visual resolution that is barely 1/100th of the printed world [Travis 1999], moderately high resolution colour displays are becoming increasingly common among all users in 1990s. As a consequence, computer interfaces that fail to use colour may not be endorsed by users although performance differences (for example, user efficiency) may not always be found between colour and monochrome screens [Travis 1990, Munns 1968, Jeffrey and Beck 1972, Cakir *et al.* 1980, Haubner and Benz 1985].

Despite the increased capability and potential advantages afforded by colour displays, the effective use of colour in designing multicolour user interfaces requires an understanding of the HCI issues, specifically the requirements of computer users during task performance.

Diaper and Sahithi [Diaper and Sahithi 1995, Sahithi and Diaper 1998] stated that there was little advice on the functional use of colour in the HCI literature. The research presented in this thesis argues that colour is an under exploited user interface (see Chapter1) feature in that it is infrequently used as a means of conveying information to users. Furthermore, there is no consistent, functional use of colour across software applications. In this context the literature on colour is reviewed. Following some background information on colour, this chapter discusses currently available colour "products", for example principles, standards (ISO 9241; Part 8), guidelines and style guides regarding the applications of colour in the design of human-computer interfaces.

2.2 The Background of Colour

The physiological psychologist Richard Boynton [Boynton 1978] says:

> "From early childhood we are easily able to recognise a property of objects, usually associated with their surfaces, that we call colour. No child, and relatively few adults, will doubt that colour is on (or sometimes in) objects."

From the earliest history to state-of-the art technology, colour has proved a potent communication medium. Artists speak of the mystical qualities of colours; people see the world as a rich tapestry of colours of different forms - coloured fields, oceans, buildings, and so on. The visual appreciation of colour has occupied the minds of physicists ever since Sir Isaac Newton discovered that when white light is passed through a prism a colour spectrum is obtained. By this means colours are obtained in a pure state. Some physicists defined colour as:

> "Colours are relational properties: they are physical properties, e.g. spectral reflectance, which are perceived in a distinctive way " [Evan 1995].

> "Colours are hybrid properties: to have a specific colour is to have some intrinsic feature by virtue of which the object has the power to appear in a distinctive way " [Maund 1998].

The colour of an object as signalled by the eye-brain system invokes a perceived, i.e. psychological, aspect that may vary with the state of the object. The same object may appear in different colours under different lighting conditions. Newton's colour disk, when subjected to rapid spinning, is perceived as white as the eye is unable to distinguish individual colours emitted by the disk. Many different colour disks can all lead to the same perception of whiteness. Colour, thus can be considered as not intrinsic to the object, but can be regarded only as a *secondary sensation* because it is not locked to a unique set of real world properties.

The major problem with colour has to do with fitting what people seem to know about colours into what science, particularly physics, seems to require of physical objects and their qualities. It is this problem that historically has led physicists, philosophers and psychologists to provide different proposals for how colour should be conceived [Maund 1998].

From the psychophysical point of view, colour can be defined as [Wyszecki and Stiles 1982] :

> "that characteristic of a visible [colour stimulus]...by which an observer may distinguish differences between two structure-free fields of view of the same size and shape, such as may be caused by differences in the spectral composition of the radiant power concerned in the observation".

The psychological dimension is very important as colours can draw attention sometimes more readily than other visual cues for several reasons. These reasons include the subjective associations of the colour, for example red for danger, and peoples' natural tendency to look for colours in the work environment.

From an anthropologist point of view:

> "Colours are socially or culturally constructed properties. To be red is for an object to satisfy such criteria that make it worthy of having the predicate 'red' applied" [Brake 1993].

Indeed, red does have a cultural meaning in UID [Diaper and Sahithi 1995]. It is possible that other colours do mean something for computer users. On one occasion the author had an opportunity to observe a twelve and a fourteen-year old boy playing some computer games. All of a sudden the system crashed and the boys could only see a blank blue screen. When the twelve year old was puzzled, the other told him that "the blank blue screen means that the computer is dead and we need to restart the system". Perhaps other computer users might also attribute the same meaning to blue on their displays, i.e. something is dead.

Colour is described using the three terms: hue; lightness (or brightness); and saturation. Hue refers to the chromatic component of the colour, the quantity that distinguishes between colours such as red, green and blue. Lightness / brightness refers to how dark or light a colour is (see section 2.2.1). Finally, saturation describes the purity of a colour, which ranges from neutral grey to pure colour.

Designers who are concerned with coding psychological dimensions appear to prefer colour coding the screen using the perceptual dimensions of hue, saturation and brightness. Travis [Travis 1991] claims that hue is more suited to formatting unrelated types of information (e.g. a spread sheet with credits and debits), and saturation to menu driven interfaces that have a number of choices that become unavailable when a particular option is selected, for example cut and paste [Macintyre 1991].

2.2.1 Colour Models

Manufacturers' representatives claim that users of computer systems appear to be excited about viewing and using millions of colours available on their lap tops and desk tops. In spite of this, however, not many colours are remembered or recognised by people. Psychologists and colour researchers [Berlin and Kay 1999] have shown that there may be only eleven categories of basic colour sensations, each associated with a well-learned name and possibly a unique physiological substrate. The prevailing doctrine of American linguists and anthropologists found a universal list of eleven categories, with greatest agreement amongst users [Smallman and Boynton 1994]. These are black, white, red, green, yellow, blue, brown, purple, pink, orange, and grey. Later experiments, however, have shown that people can distinguish many more colours than this list of eleven. Titchener, an experimental psychologist, demonstrated that there were about 35,000 elementary colour sensations of the human mind [Travis 1991]. Results of this kind demonstrated the inadequacy of relying on a simple naming system for classifying colours [Kraupner-Stadler 1991]. In this context, colour researchers have developed methods of colour specification, in particular, from the point of view of generating colour on computers and writing computer applications. Colour perception is three dimensional, therefore, any specification system needs to be based on three primaries and can then describe any colour desired by the right 3-tuple [Hill 1990].

Over several decades, many colour models/spaces have been developed [Macintyre 1991, Hope and Walch 1990, Jackson *et al.* 1994, MacDonald and Vince 1994] which vary in how colours are specified. For computer applications, often the dominant colour model is the RGB (Red, Green, Blue) one. This specifies colour as an additive combination of the Red, Green and Blue primaries. It is in widespread use owing to its deliberate close correspondence to colour CRT hardware. Two other hardware- oriented colour models are YIQ (Y is the luminance signal; I and Q are hue signals) and CMY (cyan, magenta, yellow). The YIQ model is used in U.S. commercial colour-TV broadcasting and is therefore closely related to colour raster graphics. YIQ involves recording RGB to improve transmission efficiency and downward compatibility with black and white television. YIQ values are constructed as linear transformations of the R, G, B light intensity signals. The signal is transmitted using the National Television System Committee (NTSC) standard. Cyan, magenta and yellow are the complements of red, green and blue, respectively. When used as filters to subtract colour from white light, they are called subtractive primaries. This model is used for certain colour-printing devices [Foley *et al.* 1997].

Unfortunately, none of these models is particularly easy to use, because they do not relate directly to colour notions of hue, saturation, and brightness. As a result, another class of models has been developed with ease of use as a goal: Hue (H), Saturation (S), Value (V) model [Smith 1978]. This is actually a deformation of the RGB model, and is also known as the HSB model, where B stands for the Brightness. Another model that attempts to provide a method of colour specification is the Hue, Lightness and Saturation (HLS) model. Like the HSV model, HLS is a transformation of RGB. HLS offers a method of colour selection slightly different than HSV and is used by some artists and display manufacturers [Angel 1997].

Foley *et al.* [Foley *et al.* 1997] describe the technical difference between lightness and brightness: Lightness embodies the achromatic notion of perceived intensity of a reflecting object. Brightness is used instead of lightness to refer to the perceived intensity of a self-luminous (i.e. light emitting rather than reflecting) object, such as a light bulb, the sun, or a CRT. Value is a carefully calibrated uniform lightness scale used in the Munsell colour system. While these terms are used in colour models such as HLS, HSV or HSB, one should be aware that they are used in these systems to represent fairly arbitrary measures of colour luminance [Macintyre 1991].

Gerristen proposed a new model which he refers to as the Colour(C)-Perception(P)-Space(S) model [Gerritsen 1989, Gerritsen 1988]. The idea behind Gerritsen's model, regardless of some shortcomings in its details, seems a sound one. It is often desirable to specify something akin to the perceived brightness of a colour when making colour selections, such as when two or more colours of the same perceived brightness are needed. Gerritsen's CPS model allows this, whereas the RGB, HSV and HLS models do not, except via experimentation. Another useful uniform colour space was created by the Optical Society of America (OSA) Committee on Uniform Colour Scales. The organisation of this space is based on Wyszecki's uniform colour space, which is in turn derived from the Munsell space [Wyszecki 1982]. It has been used in art, for example by artist Karl Gerstner. These new models are based on original colour models developed by artists and scientists such as Chevreul [Chevreul 1967], Munsell [Munsell 1947] and Ostwald [Ostwald 1931].

Currently, the colour model that has become basis for all industrial colorimetry is the one developed by the Commission Internationale de l'Éclairage (CIE). Three reasons are given for the popularity of this model: (i) it is based on psychophysical studies; (ii) it is thoroughly documented; and (iii) it has the potential to be the basis of an international standard. CIE defines three special supersaturated primaries: X, Y and Z, which replace the red, green and blue, and correspond to imaginary primary light resources. This model however offers a precise means of specifying colour stimuli as positive combinations under fixed viewing conditions and provides an excellent reference for industrial colour measurement [Jackson *et al.* 1994, Hill 1990].

Hunt [Hunt 1991, Hunt 1995] has proposed another colour model based on extensive psychometric experiments. The model uses tri-stimulus values (XYZ) of a colour and calculates the predicted appearance attributes of lightness, colourfulness and hue (LCH), taking into account a range of viewing parameters, for example light sources, luminance levels, surrounds (white, grey etc.) and media (monitor, print etc.). Jackson *et al.* [Jackson *et al.* 1994] describe this as the most comprehensive computer-model of human colour vision to date.

Further, technological advancements in media have initiated the need for some more colour models/spaces. As the Web medium matures users started looking at it with a more critical eye, seeing not just the possibilities, but also the limitations. Both content-providers and Web surfers are expecting more: higher quality images, faster download times and accurate colour. Every business that has entered into the on-line realm has encountered its own set of challenges, yet underlying every venture is one common feature - the need for Web pages to be viewed as they were designed. In particular, producing accurate colour that reflects their original design across different media is often important because corporations have invested millions of dollars in their logo and brand identity, making it critical for the colours representing them to appear uniformly and accurately wherever they are used. This further complicates colour management over the Internet because each machine reads, displays and prints colour in its own characteristic way. The emerging standards designed to handle colour fidelity between varying devices therefore need to address requirements from the Internet.

Recently, the International Colour Consortium (ICC) has proposed solutions to problems in communicating colour in open systems and on the Internet. The ICC profile format however does not provide a complete solution for all situations. A common standard RGB colour space addresses these issues. In this connection, Hewlett-Packard and Microsoft propose the addition of support for a standard colour space sRGB within the Microsoft operating systems, HP products, the Internet, and all other interested vendors [Stokes *et al.* 1996]. The aim of this colour space is to complement the current colour management strategies by enabling a third method of handling colour in the operating systems, device drivers and the Internet that utilises a simple and robust device independent colour definition. This will provide good quality and backward compatibility with minimum transmission and system overhead. The main intent is to promote its adoption by showing the benefits of supporting a standard colour space, and the suitability of the standard colour space, sRGB. That is, sRGB is proposed to overcome many application developer and end-user reservations to adopting colour management. One drawback of sRGB however is that it does not deliver the same level of colour accuracy or the colour gamut that a device-independent profile provides [Pantone 1999].

All these colour models were created for specific purposes and are appropriate only for certain applications. There is widespread belief, and some evidence, that different colour models produce significant differences in human performance. Consequently, they were also subjected to criticism and extensive evaluation. For example, Douglas and Kirkpatrick [Douglas and Kirkpatrick 1996] performed colour-matching experiments and found no differences in speed or accuracy between the RGB and HSV colour models, but found that increasing feedback improved the accuracy of matching. The feedback of experiments of similar kind may be an important factor in usability of designing a colour selection interface. Methods for converting colours described in one colour system into those of another system have been proposed, so that one can convert, for example from CIE co-ordinates into familiar RGB co-ordinates. An understanding of some basic aspects of these colour models is helpful for UI designers to use different colour features in designing colour user interfaces.

There are theories explaining many aspects of colour perception, but still there is no agreement on a comprehensive model of human colour vision. Colour researchers claim that it is necessary to develop a most comprehensive colour model and, colour being a secondary sensation (see section 2.2), this model must be psychological.

2.2.2 Colour Usage

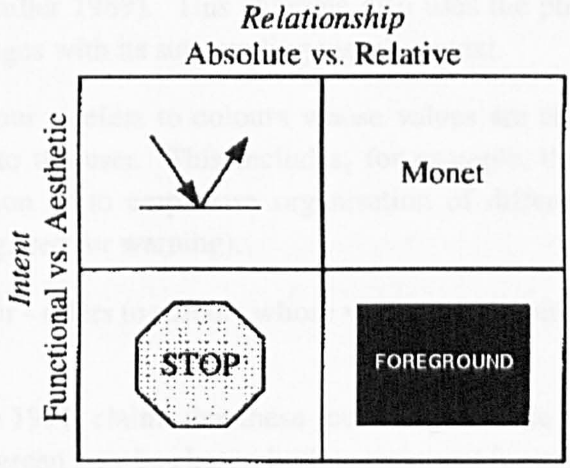


Figure 2.1: Colour usage distinguishes between colour relationships and colour intent. Typical examples include ray-tracing, the Impressionist paintings of Claude Monet, stop-sign red and adequate foreground / background contrast. Taken from Macintyre [Macintyre 1991].

Macintyre [Macintyre 1991] proposes a categorisation of use of colour. Figure 2.1 represents the usage of colour divided into four categories by distinguishing between colour relationships and colour content. Colour relationships distinguish between colours whose values are determined from an absolute specification and those whose values are determined relative to the values of other colours. Colour intent distinguishes between colours whose values are intended to serve a utilitarian function and those whose function is aesthetic.

- Absolute Colour - refers to colour values that are meaningful because of their absolute position in the colour space. For example trademark colours: the trademark blue used by Sun is different from the blue used by IBM. The exact Blue colour in each case is trademarked by its absolute colour value.

- Relative Colour - refers to colour values which are meaningful through their relationships. Examples include the relative colour techniques used by impressionist and postimpressionist painters, such as Turner, Monet and Van Gogh, to create more vibrant and intense colour impressions. By exaggerating the colour values in the shadows of a painting, for example, a viewer's perception of the colours outside of the shadows can be enhanced, possibly beyond the limits of the paint gamut. The broken-colour technique[1] developed by Monet and Renoir, may be considered as an example [Quiller 1989]. This example also uses the phenomenon that colour perception changes with its surrounding visual context.
- Functional Colour - refers to colours whose values are chosen to provide some sort of benefit to the user. This includes, for example, the use of colour to aid in comprehension or to emphasise organisation of different types of important information (e.g. red for warning).
- Aesthetic Colour - refers to colours whose values are chosen for their visual appeal.

Macintyre [Macintyre 1991] claims that these four categories are by no means exclusive. A particular shade of green may be chosen both to represent functional use and to provide sufficient contrast with another colour to enhance readability. In this case, there are both functional and relative aspects to be considered. Most colours on a display can be described by more than one of the four categories listed above. The following examples describe each cell in Figure 2.1.

- Absolute Aesthetic - refers to any colour whose aesthetic properties depend on its absolute colour value. For example, consider ray tracing - an image generation technique in which colour values are precisely computed and displayed exactly as specified. The colour values used in images serve no function aside from creating a photo realistic image (e.g. picture of an aircraft), which is largely an aesthetic issue.
- Relative Aesthetic - While some aesthetic colour choices are absolute as in realistic rendering techniques, many others are determined relative to existing colours. This is because an easy way of achieving harmonious colour is to select ones that have a close relationship to each other.

[1]Colours are painted close together on the canvas, so that the viewer's eye would blend them into a different colour.

- Absolute Functional - commonly used in interfaces to create an association between colour and a certain function. The association could be simple or complex. For example, a particular shade of blue is associated with IBM, or red with signalling emergency states or some critical information. In the first, absolute colour is a requirement and in the latter any shade of red will serve the function. Both these cases of functional colour need to be supported by the system.
- Relative Functional - used in many different ways, for example to ensure legibility or to show the logical / structural organisation of interface widgets. Ensuring legibility is an important functional requirement in UID. Studies conducted by Legge [Legge 1989] show that reading speed dropped when foreground- background contrast was reduced. Reading error counts also increased as the contrast deteriorated.

Macintyre's model provides a useful categorisation of colour usage, but applying this model to a wide range of situations may be relatively difficult. In many cases, these categories refer to implicit functions, which the designers might not even be aware of.

2.3 What should Colour be Used for?

Over several decades, studies [Christ 1975, Meier 1988, Sticker 1992] showed that colour is a more useful visual feature than size, shape or brightness in search and identification tasks[2]. Additionally, colour associations are remembered over long periods of time. For example, when two objects differing in colour are for different purposes, colour serves as an important cue to distinguish the two. These factors, combined with peoples' preference for aesthetic colour displays, form the basis for the increased use and production of colour displays of different types, for example computers, televisions, supermarket tills, cash machines and various consumer goods. In the past, researchers such as Licklider [Licklider 1979] pointed out that "information technology is flourishing .., but the field of education is not taking much advantage of the new technology". Today this situation is radically different in the technically advanced geographical areas. With sound government support, schools and colleges are now fully equipped with high performance computers thereby encouraging every student to use them in performing day-to-day activities. Consequently, there is an increasing demand for computers, in particular those with colour displays, and much is made of the use of colour in selling computer systems.

[2]The comparative problem of the 'strength of the psychological representation of equivalence' between colour and other stimuli, for example what colour is equivalent to what size, however was not established.

Supporters of colour UIs, such as Smith [Smith 1987] have made claims such as:

> "...colour-coded displays are 200% more effective than other codes such as brightness and shape ...The use of colour reduced training time, improved readability, cut search time, increased performance and is preferred to monochromatic displays."

While some caution has been expressed, this usually has been in a context of viewing the use of colour positively. For example, Tufte [Tufte 1990] says:

> "...misuse of colours increases the time needed to interpret the display, thereby decreasing productivity"

MacDonald [MacDonald 1999] agrees with Smith and Tufte when he says:

> "...Colour therefore provides an important dimension in visual communication: when used well, it can greatly enhance the effectiveness of a message, but when used badly it may substantially impair it."

Corbett and Kirakowski [Corbett and Kirakowski 1987], however, capture what is still the state today when they say:

> "...colour coded items tend to be used as a starting point for looking for patterns ...there is very little knowledge available about how best to use colour in order to facilitate high-level cognitive tasks that go beyond the seek-and-identify task so frequently studied in the past..."

2.3.1 Users Prefer Colour Displays

Studies conducted by Schutz [Schutz 1961], Christner and Ray [Christner and Ray 1961] and Tullis [Tullis 1981] have established that users prefer colour coded displays to monochrome and achromatic codes, such as blinking, reverse video and underscore. Munns [Munns 1968] concluded his article by saying:

> "...the use of colour made S's (subject's) task somewhat easier but did not increase S's effectiveness. S's report feeling more secure with the colour coding."

This preference for colours still exists today and, perhaps may be expected to continue forever [Pantone 1999]. Jeffrey and Beck [Jeffrey and Beck 1972] and Chase [Chase 1970] both used intensive questionnaire and interview techniques to obtain subjective evaluations of the effectiveness of colour. In both the studies, test users emphasised the increased naturalness of the colour displays and *their belief* that colour improved their ability to detect details, although colour did not actually improve detection or identification in the Jeffrey and Beck study. Further, a general subjective report stated that colour in displays was less monotonous and that it produced less eye strain and fatigue [Meier 1988].

Empirical studies described later in this thesis (Chapters 6 to 12) used questionnaire and interview techniques to obtain information on users' preference for colour interfaces. In the context of colour versus monochrome displays, 49 out of 50 people interviewed preferred colour displays. A majority of users seem to believe that colour coded interfaces improve their productivity by making the work environment more enjoyable to use. This positive attitude of users towards use of colour combined with the widespread use of colour displays raises some important concerns for the HCI and UID community. For example whether colour can be used as a primary cue in providing information as in the case of traffic signals. i.e. 'design for colour' instead of 'design for monochrome first' (see section 2.3.3). The London underground map could be considered as a typical example of use of colour as a primary cue. Figures 2.2 and 2.3 show this underground map in colour and in monochrome, respectively. The effectiveness of colour is obvious when one compares it to a monochrome version. A general observation of use of this map indicates that people immediately associate and identify the piccadilly line with a particular shade of blue, but not with a different line pattern as shown in Figure 2.3. Reactions of users generally provide a realistic base in designing applications [Robertson 1979]. That is, there are applications where colour can be best used as a non-redundant code in signalling important information. This domain of colour usage needs to be further explored in designing user interfaces both on the Web and traditional UI platforms.

2.3.2 Understanding of Colour is Meagre

While studies report a marked preference for colour displays over monochrome displays [Travis 1990], it appears that colour is one of the most poorly understood aspects of user

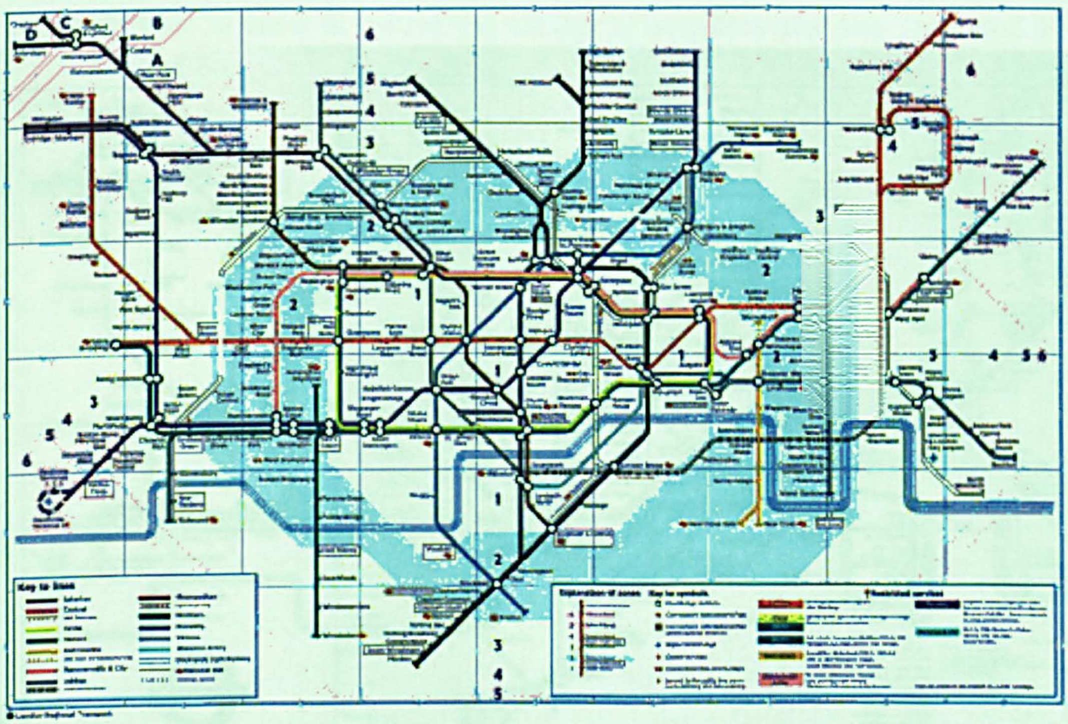


Figure 2.2: London Underground Map in Colour (©London Underground)

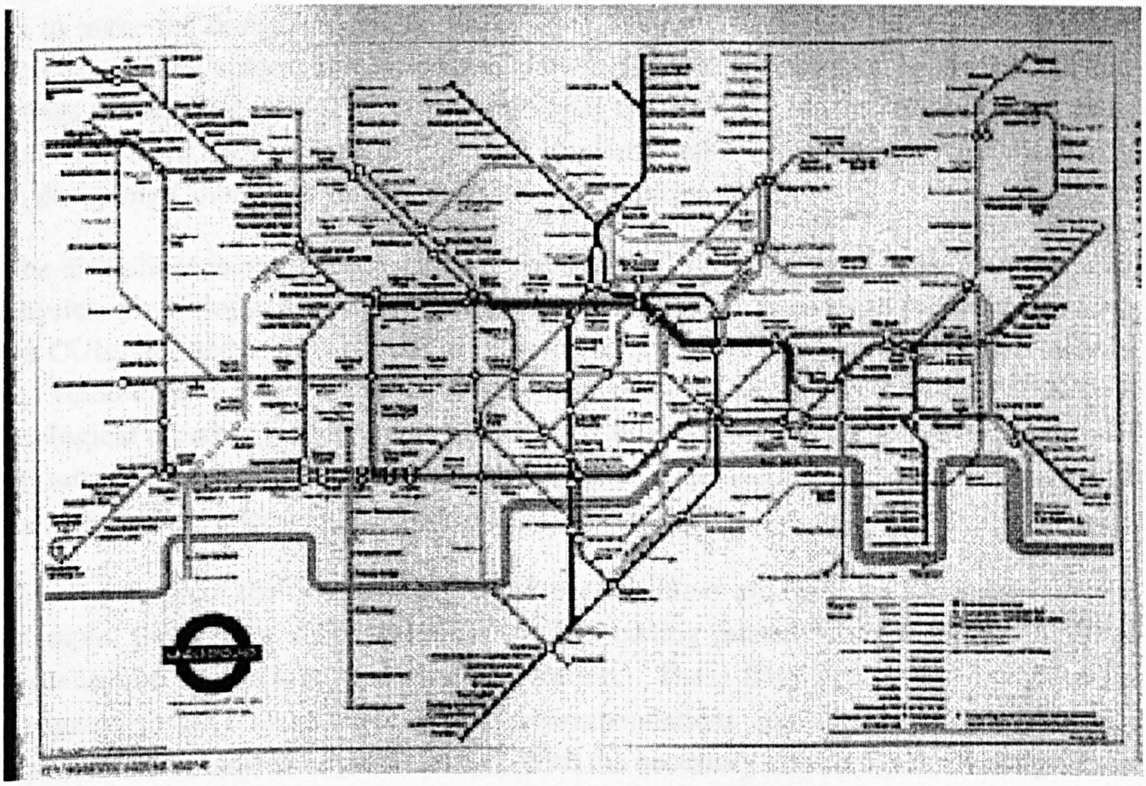


Figure 2.3: London Underground Map in Monochrome (©London Underground)

interfaces both by application designers and end users.

The paucity of understanding on the part of UI designers could be attributed to several factors, one of them being inadequate research in this area or its dissemination. Many of the available principles or guidelines on colour are too general and so are often difficult to realise in practice. In the absence of required knowledge, colours are often selected seemingly at random, with no regard for even the available design principles [Brown and Cunningham 1989]. Experience shows that developers often resort to bland or extreme colour schemes in order to reduce the number of variables that they must deal with so as to make the design task more manageable. Unfortunately, the functional benefits of colour, such as context resolution and showing logical relationships, are often the first things that are discarded in an effort to simplify the problem. Macintyre claims that many tools available are very primitive in their implementation and do not provide much help in designing Colour User Interfaces (CUIs) [Macintyre 1991].

The disciplines relevant to the study of colour and CUIs. They include computer science, physics, physiology, psychology, art and graphic design. In order to create colour tools for CUIs, selected aspects of each of these subjects need to be understood. These include: (1) video display hardware to understand how coloured images are created, (2) the psychological aspects of colour perception, (3) the effects of colour on user performance and (4) artistic models of colour and colour harmony to discover techniques of assisting with aesthetic colour selection [Macintyre 1991, Corte 1986].

Usually, many of the decisions about the use of colour are made by UI designers. For example, the way in which colour is used in coding related visual elements in a data visualisation or realistic rendering application. These days it is normal practice for designers to only make some global recommendations and allow users to customise colours to their individual preferences[3]. With the necessary knowledge, a UI designer can select a few colours and come up with several plausible colour schemes. It is therefore important to study different colour aspects simply to develop this knowledge and to learn some definite ways of developing and delivering colour schemes.

The author has conducted two sample surveys with the intent of determining the quantity

[3]While some users prefer selecting aesthetically pleasing colours, what is particularly frustrating for others is that selecting colours for the display is something that is tangential to the reason they are using the computer in the first place. A psychology student at Liverpool University was critical about selecting and changing colours of his computer when he moved from a monochrome machine to a new colour machine. Although he found the colours garish, he quickly gave up any attempt to change colours. He mentioned that either UI designers or the system administration people should have taken care of the visual aspects.

and nature of recent knowledge available on the functional aspects of colour. The first survey consisted of an on-line search of publications archived at the university of Berkley [URL 1998a] Web site and similar sites. The search yielded two hundred and eleven hits with both "colour" and "color" as the search terms. An inspection of the search results indicated that approximately 85% of the material published was concerned with issues relating to colour, such as perception and reproduction of colour for different media, and other physics and psychophysics aspects rather than its functional use. In the second survey conducted using BIDS (a library service available at Liverpool University, UK; also available at other universities), out of a sample of two hundred papers, 75% of them were concerned with the above mentioned issues. Overall, 15% of them studied and reviewed colour from the standpoint of its impact on user task performance when used in grouping and coding objects on the display. Most of the recommendations are too general leading to many, often confused, interpretations of applying them to particular design issues. Clearly, there has not emerged what might be called a **functional domain** of colour, with well-defined designs or schemes proposing creative functional uses of colour.

2.3.3 Empirical Studies

Extensive reviews and experiments conducted over two decades ago brought about mixed conclusions on the use of colour. Researchers have explored uses of colour under several categories: for example: in learning and remembering, discrimination and identification; as a redundant code; usage of colour against alphanumeric symbols and monochrome displays. Some comprehensive reviews pertaining to the use of colour in uni-dimensional and multidimensional[4] displays were published, in particular Christ's [Christ 1975]. Conclusions drawn from various experimental studies [Carter 1979, Carter and Cahill 1979, Christ 1975, Christ and Corso 1983, Robertson 1976, Robertson 1982] indicated that memory for a target colour deteriorates less during a retention interval compared to remembering target size, orientation or shape, i.e. colour associations have been shown experimentally to be learned and remembered over longer durations. In the case of discrimination and identification, colour was found superior to size, brightness, and shape in searching for and identifying items that vary in only one aspect, for example only in colour or shape. The gain in accuracy was reported to be at least 176% better than size,

[4]In Unidimensional displays, information elements differ from one another in terms of single dimension or attitude; for example colour, brightness, size, shape etc. The selected dimension for a uni-dimensional code may have many levels as are operationally feasible; however only the selected dimension may vary. All other dimensions are held constant. In multidimensional displays, a combination of number of coding dimensions can be used to convey specific information [Silverstein 1987].

32% better than brightness and 202% better than shape[5].

Colour as a redundant code, i.e. when used along with other features, such as spatial location, was found to decrease search time in symbolic displays if the user knows the colour of the targets [Markoff 1972]. In a study of highly dense aircraft situation displays, redundant colour coding significantly reduced both response time and error rate ([Kopala 1979] referred to in [Reising and Aretz 1987]). In this situation, the effect of colour can be described as cumulative. In contrast, some studies [Garner 1972] which used colour as a redundant target dimension, did not find measurable advantages with the use of colour over black-and-white displays. Garner's [Garner 1972] analysis of the results indicated that for redundant dimensions to aid identification there should be integrity of all target dimensions. Graphic design theorists such as Bertin [Bertin 1983] strongly recommend the use of colour in conjunction with other visual variables, for example size, shape, orientation, position and texture for effective design of presentations on computer displays [MacDonald 1999]. One of the interesting results was that people identified colours with lesser accuracy than alphanumeric symbols. One possible reason may be that the accuracy of identifying attributes of visual targets is directly related to the person's previous experience in responding to values within these dimensions [Macworth 1963]. Similarly, some studies [Christ 1975] have demonstrated the interference of colour with the accuracy of identifying, and the speed of locating, achromatic target attributes. This finding about the use of colour might severely limit its application in situations where colour is not the only target feature of concern, or where colour might be functionally irrelevant. These studies, however, did not provide sufficient data to explain why colour interfered with the accuracy of identifying or locating achromatic targets. Egeth and Pachella [Egeth and Pachella 1969] suggest three possible reasons: a) colour interferes with the ability to discriminate other target dimensions, b) colour distracts the observer so that he does not attend to other target dimensions, and c) colour-naming responses are more highly practised than are the naming responses for other target dimensions, so that the effect could be a result of response completion. Each of these hypotheses has strong implications for the use of colour and, more particularly, for the question of how to reduce the interference effects which may apparently accompany its use. Further, empirical studies have investigated colour versus monochrome effects. While some produced performance gain using colour over black-and-white or monochrome as large as 200% [Christ 1975], other studies did not show significant performance difference between them [Munns 1968, Jeffrey and Beck 1972].

[5]The studies however raise some interesting questions, such as those regarding the equivalence between brightness and shape.

Colour has therefore been found to be an effective factor in enhancing performance under certain conditions but detrimental under others. Research conducted in 1970s, may not be directly applicable to today's colour interfaces because most of the research was conducted on early generation CRT displays and using other. Further, it should be noted that the majority of the studies used very simple tasks (search and identify in particular) which are not representative of tasks carried out on VDUs in realistic work environments. These early research findings may therefore not be completely applicable to assessing productivity gains and losses in today's computer applications.

Studies later in the 1980s reported substantial performance improvements using colour. Green *et al.* [Green *et al.* 1987] used colour coding in tactile displays and found that when colour was used for initial classification, errors and processing time were both substantially reduced. Information portrayed by traditional shape coding required 75% more processing time and produced 200 to 800% more errors than colour coding [Sidorsky *et al.* 1984]. Sticker [Sticker 1992] confirmed previous findings: on remembering colour associations over a longer period of time than other attributes; and on the usefulness of colour in discrimination and identification. In the context of redundant colour coding, Robertson [Robertson 1982] describes that the use of colour avoids having to add more symbols or characters to complex displays. Other research has indicated that colour plays an important role in designing several applications; for example, improving the overall safety in process plants. Certain components of processing equipment are painted to comply with the 'Safety color code for marketing physical hazards', an ANSI standard. This particular code specifies using red to denote danger and stop, orange to identify dangerous parts of machinery, yellow as a cautionary indicator, green to signify safety, blue to denote non-safety related messages and equipment under repair, black on yellow combination to denote radiation, etc. Reising and Aretz [Reising and Aretz 1987]. Studies conducted by Travis [Travis 1991] describe colour as a potent visual cue when used to make an object 'pop out' from the surrounding objects. For example, finding a book with a specific title is a slow process; identifying the colour of the book is almost instantaneous Treisman and Gormican [Treisman and Gormican 1988]. In addition, the studies conclude that using colour to delineate words makes reading very easy.

To examine more recent findings, approximately 35 empirical studies conducted between 1990 and 99 have been reviewed. These studies examined different aspects of use of colour in designing visual displays including its usage as a redundant or non-redundant code. As in previous research, some recent studies have found significant performance improvements using colour [Jacko *et al.* 1999, Alonso and Norman 1998, Braun *et al.* 1994, Lang *et al.* 1994]. Studies demonstrated that colour recognition can be used to good effect in a symbolic programming environment [Batchelor and Whelan 1997, Batchelor and Whelan 1993]. Healey *et al.* [Christopher G Healey and Enns 1996] showed that colour improves user performance in rapid and accurate numerical estimation. Swan and Perlman [Swan and Perlman 1995] reported an experiment which compared the relative effectiveness of colour coding, texture coding, and no coding of target borders, to speed visual search. Results showed that colour coding improved search time compared to the other codings. Duijnhouwer and Zwaga [Duijnhouwer and Zwaga 1995] conducted studies investigating the merits of colour coding compared to that of shape, finding that search with shape coding is inferior to that with colour. Nugent [Nugent 1994] claims that colour-coded NATO (North Atlantic Treaty Organization) symbols yielded faster, more accurate performance than other symbol sets. Converse *et al.* [Converse *et al.* 1992] tested the hypothesis that colour coding can be used to enhance the speed and accuracy of performance on a focused attention task. The study was carried out with different experimental conditions including a monochrome condition. The results indicated that the colour codes significantly reduced task response time without degrading task accuracy. In contrast, studies conducted by Wickens *et al.* [Christopher D Wickens and Sarno 1995] indicated that while comparing conventional 2D graphs with 3D graphs and colour based graphs for presenting 3-dimensional data, performance with the colour display suffered badly in both speed and accuracy. Similarly, Smith *et al.*'s [Smith *et al.* 1995] findings supported a general argument that colour should not be preferred automatically, but rather its utility depends on the cognitive demands of the task for which the display is designed.

Products of recent research include expert systems, for example ACE (A Colour Expert System). The purpose of these systems is to select colours automatically for the sort of user interface found on the Apple Macintosh or Xerox STAR based desktop environments [Macintyre 1991]. Tools such as ColorWeb and ColorServe (a Java application) [Pantone 1999] have been developed to provide Web page authors with an Internet safe colour palette, which ensures that the colours appear as intended irrespective of the end user computer and browser. The tools allow interactive colour selection. Other products including methodologies for producing maximally discriminable, nameable colours in certain visual displays have been proposed [Laar *et al.* 1997].

2.3.4 Products of Colour Research

The products (referred as REPs in this thesis, see section 1.4.2) of previous research on colour include design principles, standards, guidelines, style guides, heuristics and rules of thumb.

2.3.4.1 Design Principles

Brown and Cunningham [Brown and Cunningham 1989] have offered a dozen principles for using colour: 1) Be consistent in the use of colours, 2) Seek advice of professional artists in the use of colour, 3) Use no more than five to seven colours in a display, 4) If you use red and green, place them in the centre of the display, not in the periphery, 5) Use blue for background or large areas, 6) Avoid adjacent use of two colours that differ only in the brightness of blue, 7) Do not rely on colour coding alone to distinguish small areas, 8) Do not use intense combinations of blue/yellow, red/green, red/blue or green/blue, 9) Use warm colours to bring objects closer to the viewer, 10) Use familiar colour codings appropriately, 11) Avoid subtle differences in colour (or) grey scale, and 12) For colour deficient viewers colour code background rather than text.

The same criticisms made of Smith *et al.*'s design principles (see section 1.3.1) apply to those of Brown and Cunningham [Diaper and Sahithi 1995]. For example, it is likely that principle 10 will be at times in conflict with principles 4 or 5. The greatest drawback, however, with Brown and Cunningham's list of principles is their incompleteness.

It is self evident from a list such as this that it is simply not clear what to use colour for. Principles 1 and 10 hark back to the STAR design principles. Principle 2 is the only one listed under the heading "Principles of Effective Colour Use" and, while it may be sensible advice, it provides evidence of our ignorance about the use of colour by implying that such knowledge is a "craft skill" [Long 1986b, Long and Dowell 1989b]. The skills of professional artists moreover are usually limited to traditional paper based media. As a result, a straightforward transfer of their skills to computer colour displays may not always be feasible.

Principle 9 is potentially interesting if one could decide why, under what circumstances, etc. it would be desirable to suggest objects are closer (i.e. the *functionality* of colour induced apparent depth requires specification). In the list, principle 10 is concerned with the general functional use of colour, but even here "appropriately" offers little operational advice. The remaining provide advice regarding selection of colours.

These principles or guidelines have emerged from previous research and from the designer's experience of working on colour displays. Today, there exist hundreds of guidelines or style guides (See Appendix B.1 for a sample listing of 50 of them). Some of them reiterate what has already been established, some are concerned with special needs of people, for example, emphasising 'design for colour-blind', and some paradoxically are still rooted in early days of monochrome display systems. For example, 'design for monochrome' is one guideline that appears in almost every book or review. While this guideline primarily recommends use of colour as a redundant code, its applicability in today's displays, now that monochrome is quite rare, needs to be addressed.

Different reasons were mentioned in the literature for recommending the 'design for monochrome' guideline. For example, Foley *et al.* [Foley *et al.* 1997] say:

> "A conservative approach to colour selection is to design for a monochrome display, to ensure that colour use is purely redundant. This option avoids creating problems for users whose colour vision is impaired and also means that the application can be used on a monochrome display."

The use of the monochrome guideline is generally promoted for three different reasons: a) to cater for monochrome display users, b) to tailor the displays to be suitable for colour blind people and c) to tidy up the design process by reducing the intervention of colour. While the points on the conservative use of colour are well taken as precautionary measures to avoid meaningless colour usage, the reasons given for limiting the use of colour require some explanation.

Monochrome displays are no longer common, not least, because people prefer colour displays (see section 2.3.1). Realistically, designing for 'monochrome users' is no longer a necessary guideline to be given attention to in the development of current visual displays. Regarding point (b), several papers concerning the needs of colour deficient people have been published in the recent years [Laar *et al.* 1997, Batchelor and Whelan 1997, Dalton 1997]. Research indicates that in Europe and North America less than 0.1% of people are unable to perceive hue, and overall approximately 8% of males and 0.5% of females are colour vision deficient in some way. Colour deficiency often prevents the perception of combinations of some colours (see Table 2.1). The most frequent form of colour deficiency is the inability to distinguish red and green. Others that could easily be confused by colour deficient people are cyan from white and blue from purple [9241-11 1997, Wright 1995]. If applications need to be designed for people with colour deficiencies, combinations of colours other than these need to be exploited. Alternate solutions however suggest that by ensuring sufficient brightness differences between colours, it is possible to eliminate confusion between these colours. Point (c) mainly suggests design for black-and-white first. In this context, Wright [Wright 1995] claims that her biggest challenge was to convince designers that pure black and white are not neutral and should not just be used as a normal practice in designing systems. She reasons that perhaps it is because the traditional starting point of a graphic design project was invariably a black pencil and a white paper sketch pad. These two colours i.e. black and white are so ingrained in designer's minds that it is hard to convince them that a white background, or black lettering can destroy the whole project (white background leads to visual problems with glare and flicker). Further, she describes a practical basis for this attachment to black and white; it is the easiest and cheapest from the printer's/printing point of view. Mapping colour design details of the display monitor onto a black and white design on paper has certain disadvantages and therefore indirectly encourages designers to prefer black-and-white designs. These considerations suggest that giving priority to monochrome is no longer necessarily the best strategy. While the needs of colour impaired users should be taken into consideration, they should not provide an absolute basis for exploiting the useful enhancements that colour can provide.

Preferred Designation		Possible Colour Discriminations	Incidence Population (%)	
By number of components	By type		male	Female
Trichromatism (3) (normal or colour weak)	Normal	L-D, Y-B, R-G	-	-
	Protanomaly (red weak)	L-D, Y-B, Weak R-G	1.0	0.02
	Deuteranomoly (green weak)	L-D, Y-B, Weak R-G	4.9	0.38
Dichromatism (2) (partial colour blindness)	protanopia (red blind)	L-D, Y-B	1.0	0.02
	Deuteranopia (green blind)	L-D, Y-B	1.1	0.01
	Tritanopia (blue-yellow blind)	L-D, R-G	0.002	0.001
Monochromatism (1) (total colour blindness)	Congenital Total Colour blindness (cone blindness)	L-D	0.003	0.002
* L-D = Light - Dark Y-B = Yellow - Blue R-G = Red - Green				

Table 2.1: Incidence of colour vision deficiencies for males and females (Results reported by Judd and Wyszecki during 1963, cited by Silverstein [Silverstein 1987].

Using or not using colour as a salient design feature can lead to different design options and the latter case forms the basis for the argument presented in the thesis i.e. colour has been an under exploited display feature.

Years of expensive effort has been expended on designing sophisticated colour displays. What should be the purpose of this effort? Clearly, it should not be merely for promoting aesthetic appeal to sell computers. Considering that colour plays an important role in the realistic environment (red, green, amber traffic signals guide millions of people and regulate everyday traffic around the globe), its usage as a powerful and popular primary cue should be further exploited in human-computer displays.

> "Colour can enhance the visual and cognitive processing of information on display screens. For example, colour can help locate, classify, and associate images (i.e. show a relationship between information)." ([9241-11 1997], Part 8)

2.3.4.2 Consolidated Lists of Colour Guidelines and Style Guides

Previously, Smith and Mosier [Smith and Mosier 1986] and Meier [Meier 1987] and others have attempted to provide a consolidated list of colour guidelines and style guides. Among recent readings, MacDonald [MacDonald 1999] had offered a comprehensive range for using colour in UID. Appendix B.1 lists MacDonald's 39 guidelines based on biological, psychological and technological considerations. These guidelines cover different aspects such as colour vision, perception, display technology, graphic user interfaces and design principles, use of colour in advertising and information coding, visualisation and imaging. In summary, he recommends five golden rules for producing effective visual displays: 1) Take account of human visual needs and expectations, 2) Conform to the colour conventions for the application, 3) Design the screen layout considering all available visual variables, 4) Be consistent in the use of colour throughout all screens in an application, and 5) Use colour sparingly, never more than is necessary for the task.

These rules emphasise the UID philosophy, importance of consistency in designing applications, and address the need to exploit all available visual display variables; the three important issues that form the basis of this research. While demonstrating uses of colour, MacDonald describes two types of colour coding: a) nominal and b) ordinal. In the first, unique colour codes are assigned to the different components or states of a system. This coding neither indicates differences in value nor an order of priority among the components. A good example of this colour coding is the London underground colour map (see Figure 2.2). In the latter, a graded and explicit sequence of colours is used to represent the value of one or more variables. While doing so usage of natural or application-related associations could be exploited, for example the hues of the rainbow form the familiar spectral sequence from red to violet and changing colours of a radiating body i.e. with increasing temperature from black through red to yellow, white and blue. In both schemes, providing a legend alongside the display would aid interpretation.

Apart from MacDonald's guidelines, there exists hundreds of guidelines and style guides concerning use of colour in designing user interfaces. Nielsen [Nielsen 1993] has made a list of UI guidelines organised within six functional areas of user-system interaction: Data Entry (199); Data Display (298); Sequence Control (184); User Guidance (110); Data Transmission (83); and Data Protection (70).

Appendix B.1 lists a sample of around fifty guidelines. It has been emphasised [Robertson 1979] that a majority of this work on guidelines or style guides has not been concerned with application trials or the testing of various methods of employing colours in running applications. Some of the recommendations therefore could be subjected to a high degree of scrutiny and need to be modified in the light of empirical evidence. Lalomia and Happ [Lalomia and Happ 1987] describe how available guidelines provide direction but not the detail required for application programmers/ designers. It is also the case that some of them provide confusing advice. For example, colour psychologist Wright [Wright 1995] describes how certain terms cause confusion to designers. For example, the phrase 'primary colours' means different colours in different contexts. Pigment primaries are red, blue and yellow (i.e. mixing blue and yellow produces green) but the primary colours when using light, for example in colour televisions and monitors, are red, blue and green; yellow is a secondary light colour. Similarly, some guidelines use the term 'strong colours' but fail to define and provide examples of what they mean as their interpretations differ across cultures and/or geographical regions.

In his foreword, Chris McManus [McManus 1995] claims:

> "What is surprising ... is how little psychologists have been concerned with the effects of colour upon behaviour. Of course physicists, biochemists, physiologists and neuroscientists now have an exquisitely sophisticated theory of how we see colour, but there the theories stop. Few colour scientists seem to have any theory of what colour is really for and how it affects people beyond its mere perception."

Current collections of guidelines on colour have far too many rules to follow in different contexts and could be seen as intimidating by application developers. Programmers and designers have difficulty understanding and applying the dozens of rules published and such lists are often considered inadequate. Reports on guideline or style guide compliance failures reported by Löwgren and Laurén [Löwgren and Laurén 1993] establish a strong basis for such difficulty. In addition, an examination of available guidelines indicates that while much work has been put into attempting to evaluate the use of colour in grouping and coding, the functional aspects of colour have not been explored in comparable depth. Previous research, despite its volume, fails to deliver functional colour designs that can be used effectively by application designers.

It is necessary that colour research should focus on delivering design solutions that could be easily understood and adopted by programmers and user interface designers. The research presented in this thesis follows the proposal of Diaper and Sahithi [Diaper and Sahithi 1995] that prescriptive methods of use of colour driven by delivery requirements, such as simplicity and consistency, should be the primary product of HCI research on colour.

2.3.4.3 International Standards

There are international standards in relation to the use of colour in computer displays. ISO 9241 deals with the specifications for displayed colours to ensure their visibility, identification, and discrimination on most computer platforms and provides important advice based on state-of-the-art knowledge of colour [Jackson *et al.* 1994]. Part 8 of the standard describes minimum ergonomic requirements and recommendations to be applied to colours assigned to text and graphic applications and images in which colours are discretely assigned. The recommendations, however, are acknowledged to be sub-optimal for users with colour vision deficiencies and include: optimise discrimination and identification of colours - systems and applications should use an achromatic background behind foreground colours or achromatic foreground behind background colours; minimise the number of colours that should be displayed simultaneously. Other parts provide many other recommendations, for example using colour as a redundant feature with some other feature, such as symbology. Colour is described as a good auxiliary code and its use in this manner is strongly recommended, i.e. *do not code only in colour.* ISO 9241 is meant to be used frequently, required by those with an interest in the use of colour in computer applications. The widespread availability of these standards to an average designer however is not so straightforward. Currently, the standards are not available free of cost to the designers. Considering the vast number of people designing colour interfaces, especially on the Web, keeping a high price on standards will discourage many designers from acquiring and consequently following them. If standards are to be followed widely, then they must be made readily available perhaps like yellow pages to every designer.

In summary: colour, used properly, is generally regarded as "a good thing", but at present there is a little or no agreement about the use of colour within and across software packages, even when they are running on the same hardware and software platforms. Grouping can be done in other ways, for example by spatial location, and emergency states can be signaled, for example, by monochrome flashing. Furthermore, the style guide compliance failures [Löwgren and Laurén 1993] suggest that delivering even more guidelines in large style guide formats is likely to be ineffective, even for professional UI designers. Their critiquing expert system approach to supporting style guide compliance thus deserves serious consideration.

One hope of the redesign (see section A.1) approach is that conventions will emerge as computer systems evolve. Such evolution is not a smooth progression and there is considerable technological pull as, for example, computer capabilities continue to increase. These conventions need not be mimicry. For example, there are some icons that have an identical, functional meaning independently of their precise visual appearance (e.g. the trash can or waste-basket for delete operations; the busy bee or egg-timer to indicate system load). Similarly, a variety of scroll bar conventions does not prevent considerable positive transfer of user expertise between systems. Once the conventions are established, they become de facto standards.

The issues raised in this thesis are directed towards HCI for failing to provide an adequate theoretical base that could be used to address the question: 'What should colour be used for?' Given the vital importance of functionality [Hammer 1984, Diaper and Addison 1992a] this issue requires serious attention of UI research community.

2.4 Designing Colour UIs

There are many perspectives on the purpose of designing a User Interface. The approach taken in this thesis is based on a task orientated philosophy [Diaper and Addison 1991] and proposes that one major purpose of designing a UI is to support end users during task performance.

People use computers to accomplish tasks and a well-designed user interface should support all possible types of user actions during task performance. LURE [Diaper and Sahithi 1995] is an acronym used to exemplify a high level view of what users want to do in most computer supported tasks:

- Learn
  Presented with novel software, etc. users need to learn how to accomplish what they wish. Such learning may vary from highly formalised training to creative problem solving.

- Use
  Usability is rightly seen as a central feature of HCI.

- Remember
  Commonly users will want to use software repeatedly. Often there is a considerable interval between using a software package and thus it is crucial that a UI helps users remember how to use software to do a task.

- Extend
  Smith *et al.* [Smith *et al.* 1982], describing the development of the STAR system, suggest:

  > "... that users will *intuit* things to do" if "There is sufficient *uniformity* in the system ... allowing lessons learned in one area to apply to others".

  How such transfer of training occurs may vary from a rational, explicit, problem solving approach to inspired guessing[6].

The LURE description obviously forms a high level task cycle as learning is never done in a mental vacuum and thus, at a minimum, is always based on extending knowledge from previous experience to the current one.

It is not denied that there are other views and there may be other, equally pertinent factors not covered by LURE. The purpose here is not to propose an exhaustive list, but to illustrate the types of behaviour that UID should support.

2.4.1 Novel Functional Design

Foley *et al.* [Foley *et al.* 1997] claim:

[6]Note: Diaper [Diaper 1989] suggests that a realistic view of the cognitive capabilities of the human mind must recognise the difference between human performance, which is superb, from our understanding of it, which is virtually non-existent.

> "Any decorative use of colour should be subservient to the functional use, so that the colour can not be misinterpreted as having an underlying meaning."

Design is a creative process and one goal of this thesis is to suggest a role for colour that could be used more systematically in designing human-computer interfaces. The design part involved thinking of what additional information users might find useful[7] that was not already generally available on most system platforms, such as the Web or traditional UIs. This problem is set explicitly in the design process; how it is solved is described in detail later (Chapters 3 and 4). In the example designs, coloured background and border features are used in the design because little concrete research data is currently available as to how these features should be used. Travis's proposal (see Figure 2.4) is an exception in that the interface design uses background colours to indicate two types of inputs: Green background indicates compulsory input and Blue, the optional input [Travis 1991].

2.4.2 Goal: Extraction of Information at a glance

A growing number of people spend most of their days working with computer displays and rapid extraction of information is an important issue in many application domains. In such situations, the UI design should be of serious concern in order to support users during task performance. Galitz [Galitz 1980] reported the results of a study which projected that a single computer system for an insurance company would process 4.8 million screens per year. Tullis [Tullis 1983] reported analogous estimates for a Bell System computer system: the Automated Repair Service Bureau (ARSB) indicate that employees using the system must extract information from 344 million distinct frames (screens, print-outs) per year. The importance of designing these frames effectively becomes clear when one considers that if the ARSB users took only 1 sec longer to extract information from each frame, an additional 55 person-years would be required just to accommodate this. Tullis raises a question, 'whether it is possible to design such a display that actually makes a difference to how rapidly the user extracts information from it'. From the viewpoint of time critical applications, Tullis's question requires some attention. A variety of techniques can be used to aid information display such as grouping related data etc. The previous studies provided several guidelines for designing different types of displays. In a complex display

[7]The concept of 'useful information' is greatly dependent on the tasks to be performed by the users. In this research, it has been applied loosely to indicate information that users might find advantageous during task performance. For example, information concerning screen areas that are selectable and non-selectable in a Web page.

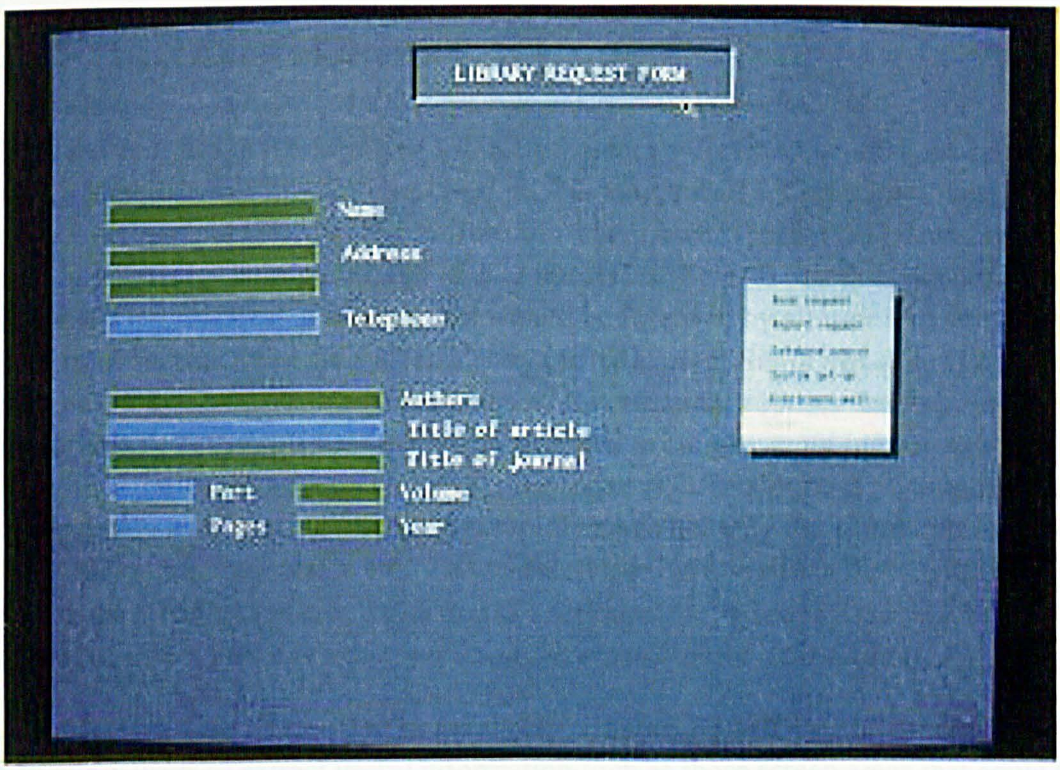


Figure 2.4: This interface to a library request system uses a green background to signify compulsory input and a blue background to signify optional output (taken from Travis [Travis 1991]).

the extraction takes much more time as the number of items in each display is typically in the range of 80 to 250 items [Ringel and Hammer 1964], especially when the search is linear. Central to the issue is that, the more the required information is presented *at a glance* (section 2.4.3), the faster is the performance of tasks.

2.4.3 User Centred Design

User-centred design has been the subject of many conferences, workshops and books in itself [Shneiderman 1992, Travis 1991]. The ten general UI principles (see Table 2.2) formulated and published by British Telecom (taken from [Travis 1991]) have been adopted by many developers in the design of user interfaces. Several practitioners describe them as the top ten practical guidelines that should be followed by all user interface designers. Principles such as these are self- evidently the children of Smith's *et al* design principles. With an emphasis on the *at a glance* concept, this research proposes that lists similar to that of the British Telecom's list could be augmented to include *rapid information extraction* as another important principle for designing displays. Users' time is important and costly and any help in extracting information rapidly saves not only the search time but reduces their frustration, impatience etc. The research therefore suggests that UI designers may need to consider information extraction as a serious issue and design the interfaces to *allow the users, where possible and appropriate, to extract useful information 'at a glance'*.

In the designs offered later in this thesis (see Chapters 3 and 4) the primary design goal is to allow users to extract useful information such as those screen areas that could take a user input when targeted by the cursor (i.e. when they have input focus) and those with hidden options (see Figure 3.1), *at a glance.*

2.5 Technical Limitations

Unfortunately, the current window systems are not sophisticated enough to let users take full advantage of the benefits of colour. The tools provided for manipulating colour are surprisingly primitive. This section describes the technical problems commonly reported while manipulating colour values of interface objects.

Ten ways to keep the user friendly
1. Use the user's model.
2. Introduce through experience.
3. Design the system to be user centered.
4. Allow the user to observe and control the system.
5. Log activities.
6. Ensure consistency and uniformity.
7. Provide immediate feedback.
8. Avoid the unexpected.
9. Validate entered data where possible.
10. Provide help information.

Table 2.2: British Telecom's Ten general principles to adopt in the design of the user interface (©British Telecom).

2.5.1 Bottlenecks in Implementation

Colour is a property of an interface component, such as a window, an icon, etc. The component is assigned a colour value at the time of creation. A typical failing of current systems is that once the component is created with a specified initial value, it is not possible to change it again during run time[8]. This is analogous to specifying an initial position for a window and not being able to move it later. The very fact that this restriction exists, whereas the inability to reposition or resize windows would be unacceptable, is evidence of the under exploitation of the functional role of colour. Many applications allow windows to be repositioned but none of them allow changing colours at run time. It is technically difficult for the designer to change the colour values of interface components as the environment changes. For example, consider a design situation where a push button is needed to be displayed initially with a blue background and then green following a simple mouse click or a small time interval. Technically, it is not feasible to implement such a design. Once the push button is created, it is permanently displayed with the initial blue value. The push button cannot be programmed to change its colour to green during the running of the application. UI tools, such as UIM/X allow the run time changing of colours only in the testing mode i.e. prior to code generation and compilation. Once

[8]For example the current Microsoft Windows does not allow changing colour values of individual widgets in an application during run time. A colour change globally affects all the widgets

the interface code is generated and compiled, the tool does not support changing colour values.

2.5.2 Resource Editing

It is generally the case that a majority of client programs on Unix and PC Windows platforms do not contain functions that allow the user to edit colour specifications. Even if they allow this, the user must edit the colour values in the resource files explicitly. Otherwise, the user must typically perform modifications in source code.

Other window systems, such as the one found on the Macintosh, allow the user to change the colour of existing windows but the tools for colour selection are still not as sophisticated as those for window positioning. In the case of the Macintosh, for example, windows are positioned for the user the first time they are created in a random fashion, spreading them out over the screen. Colours, on the other hand, are always the same by default, providing not even the slightest variation to what is specified by the user initially [Macintyre 1991].

Tools, for example 'Editres' on Unix platforms, allow users and application developers to view the full widget (interface component) hierarchy and help the users to construct resource specifications, apply the resource to the application, and view the results dynamically i.e. when the application is running, provided that the X toolkit client complies with the Editres protocol. Unfortunately, only a small number of applications comply with resource editor protocols, such as Editres. These editors allow the user to modify the colour properties such that, at any time, the interface component can hold a single value. As mentioned in previous section, it is technically not feasible to create a UI component, such as a push button, that changes its colour value more than once during run time (for example, from yellow to white). A comprehensive colour management system, that allows the changing of colours dynamically during runtime, would thus be useful for developers who use colour in designing interfaces.

2.6 Summary

In general, colour can be used on visual displays for the following reasons:

1. First, it may be used for aesthetic purposes. What matters for these applications is simply that the final display looks appealing. To achieve this, the rules of colour harmony may be followed.

2. Colour may be used on displays to represent realism. This includes applications such as computer aided design, where products are built and may be shown to customers as an example of the finished version; for example in the printing industry, where precise colour specification and colour judgements must be made.

3. Colour may be used for formatting purposes: that is, to segregate or group different types of visual information. Further, addition of colour contrast can increase symbol visibility and reduce display brightness requirements.

4. Colour may be used for coding: that is, symbolically, to signify meaning (e.g. red indicating danger) or to serve a purpose beneficial to the user in performing tasks (see Chapter 4).

Users of computer documents do not just look at information, they interact with it in novel ways that have no precedents in paper document design [Edwards and Hardman 1989]. Excellence in interface design, i.e. designing how the user is able to access the electronic information, makes the documents useful. Consequently, an important design consideration in any UI project should be to analyse the relationship between the visual and functional goals, and delivery and product requirements (see section 1.4.2).

Software designers have not explored using colour in UI design in great depth. It is therefore likely that many new useful ways of applying colour are still to be invented. The key element that is missing from most existing computer systems is that there is little functionality attached to visual features, such as colour, on user interfaces. By attaching functional meanings to colours, the display system should be able to assist the user in performing the tasks. Functional use of colour has been effective in other areas (traffic lights, underground maps and safety colour coding are some excellent examples). It is at least possible therefore that functional colour use would make an important contribution to enhancing user performance in computer based tasks.

Previous studies have not made significant attempts to exploit the potential functionality of colour in designing human-computer interfaces. The work reported in this thesis aims at contributing some results in this direction.

Chapter 3

Pilot UI Colour Design

3.1 Foreword

This thesis is investigating the use of colour as a vehicle to transfer information, *at a glance*, from the system to the user, while exploring the REP (Research Engineering Product, see section 1.4.2) design of appropriate colour schemes and the delivery requirements that drive the REP design. The research identifies consistency as an important property of the product as well as the principal delivery requirement, and attempts to design methods that ensure consistent visual navigational structures using colour.

Sections 1.3.2 and 1.3.3 form the basic premise of this chapter. That is, functionality being the driving factor, this research attempts to develop solutions that demonstrate consistency and standardisation also. These are closely related to both traditional UI and Web design and to implementation problems that users have been experiencing to date. Some feasible approaches to providing solutions to some of these problems may include delivering a set of creative, functional, consistent and simple designs to people, in particular to designers who are not HCI experts. This chapter presents the initial product of this research: an example high-level colour design for traditional UIs. The colour design is discussed with respect to proposed functionality and consistent usage both within and across applications (and platforms). The effect of the colour design on user task performance is demonstrated later in Chapters 6 to 8.

3.2 High-level Colour Design for Traditional UIs

3.2.1 Visual Interface Design Features

The visual component of user interfaces comprises display features like menus, icons, colour, layout and transient events (for example, flashing, on moving objects) etc. [Robertson 1979, Lansdale and Ormerod 1994, Travis 1991]. Considerable effort has been employed, both theoretically and in practice, on UI features such as menus and icons but their consistent use is often considered secondary. Consider for example, the selection criteria adopted in menu design across Unix and Microsoft Windows platforms. The Unix window platform allows users to select menu options using the "click-hold and drag-release" method whereas Microsoft employs a different method "click-release-move-click". Human factors research has not attempted to find a performance difference between the two menu designs but their inconsistent use across platforms has been shown to cause both user frustration and poor performance [Whiteside *et al.* 1988]. A second example is that company Web sites use icons to represent both their corporate logos (non-selectable) and operations or functions that are selectable. Another example is the most frequently quoted Macintosh menu bar which consists of a number of pull-down menus with names as well as a single menu to the far left marked by an Apple logo [Nielsen 1989, Tognazzini 1989].

Inconsistencies such as these can impair user performance. As Nielsen [Nielsen 1989] says "these small inconsistencies are no big problem, but however minute the effect of inconsistency may be, it may cause some sort of frustration and affects easy learning and, in turn, usage." Consistency problems can occur across applications and also within whole product families of software.

Inconsistencies may arise for a number of reasons associated with the development process. These include inadequate specification of design objectives, failure to understand user requirements, misidentification of common functionality, misclassification of generic design modules, and inadequate design specification and many other organisational factors such as distributed design and development groups [Neale 1993]. Yet another example of inconsistency related to the design process i.e. due to differences between the designer model and user model, would be the file copy action in Windows 95 (as well as its predecessors and successors). The copy action displays moving icons, i.e. each icon representing the physical object "paper", between two folder icons. The same paper icons are displayed even when the copy involves audio files, causing possible confusion to users who may associate the paper icon with text and/or graphics and are not sure that the same icon also represents audio content. This is a typical example of metaphorical inconsistency [Braudes 1991]. The standard audio symbol such as the speaker icon that Microsoft already uses on its platforms would have been a better symbol in this particular situation. A similar case can be made for transferring video files.

3.2.2 Functional and Consistent Use of Colour: Pilot Product

As described in Chapter 1 there is no shortage of available UI display features but some of them, for example colour, appear to have been under exploited. The remainder of the chapter concentrates on the novel functional use of colour in UIs and illustrates the pilot product of this research. The purpose of the particular colour design offered below is to illustrate the possibility of a novel functional allocation of colour for the purpose of achieving one type of consistency (i.e. semantic) across applications. The aspect of functionality this research intends to use colour to convey is the nature of the information displayed in, and the input acceptable at, the various regions of a GUI.

Lewis C.H. *et al* [Gould and Lewis 1985] stated (quoted in Nielsen [Nielsen 1993]):

> "Consistency is one of the most basic usability principles. If users know that same command or same action will always have the same effect, they will feel more confident in using the system . . . "

The proposed use of colour below is intended to support user behaviour during task performance across applications [Diaper 1987, Diaper and Addison 1992a]. Primarily, the design allows users to tell *at a glance* those screen areas that could take a user input when targeted by the cursor (i.e. when they have input focus). Secondly, the design proposes that colours will indicate the type of user input that is possible. Thirdly, the proposed colours provide consistent semantic information both within and across applications. In other words, the same information is presented in the same way, using colour, to facilitate easy learning and easy recognition, regardless of the application being used.

The design binds background colour to UI input style thus:

Background Colour	*UI input type*
White	Text
Blue	Nothing
Red	Anything *For urgent user information only*
Yellow	Mouse
Green	Hidden options

The selection of UI input types might be considered general. One main reason behind this selection is that the design intends to use very simple and commonly used input categories across different platforms. On low level design details, while the proposed designs refer to the background colour of a screen display area, an alternative might be to use only a coloured border. It is predicted that users will enjoy positive transfer of training from systems that used coloured backgrounds to those that used coloured borders, provided the design conventions were adhered to across the systems. With the proposed background colours, the textual foreground colour will be chosen for its readability, aesthetic appeal, etc. and will therefore be black by default.

3.2.3 Rationale for Choosing Blue-Green-Yellow-Red-White

The colours used in the proposed design are derived from the basic colour list [Travis 1991, Wright 1995] (white, grey, black, green, yellow, blue, red, pink, brown, orange and purple) because these colours are easily remembered and distinguishable from each other by the vast majority of users. Psychologists have shown that people are happy to classify virtually all colours by using these basic colour names [Travis 1991]. Regarding the number of colours: studies have shown that coding in applications rarely need more than seven and often only four colours [Robertson 1979, Chapanis 1965]. This research has considered only a maximum of five colours in proposing the designs. One possible way to reduce the difficulty of selecting a specific set of five colours would be to reduce the size of the basic colour set by only using combinations of RGB (see Chapter 2) display primaries i.e. Blue, Green, Yellow, Red and White. The rationale for choosing these particular bindings is briefly described below:

- White – While white background can lead to visual problems with glare and flicker, optimum readability has generally been reported for black text on a white background, thus following Rivlin *et al.*'s [Rivlin *et al.* 1990] screen design guideline to:

> "Choose a text colour/background colour combination which maintains a high contrast between the letter and the background."

- Red – Following existing convention the design reserves red for emergency information, warnings and so forth. It is assumed that other features of a UID for such critical information will inform users of what they must, should or might do and therefore this colour will be unconstrained as to user input type.

- Blue –

  The human blue colour visual system is complicated and known to be poor for pattern recognition [Mollon 1982] and is the reason for Brown and Cunningham's principles 5 and 6 (see section 2.3). As a result, blue is used to inform users of screen features that do not take any user input. Since such UI components do nothing apart from provide information it seems reasonable to use blue for this purpose because it is the worst colour for human visual pattern perception.

- Green and Yellow –

  Colour guidelines indicate that shifts in hue is good for showing categorical (nominal scale) distinctions (same or different). For example, London underground map (Figure 2.2). Based on this, the current research uses green and yellow, which differ in hue, for indicating different input categories [URL 2001a, URL 2001b]. Further, guidelines also suggest that green and yellow are generally known to be used to capture attention (e.g. traffic lights) [Nijjar 1976, Sullivan 1997].

3.2.4 Examples of Colour Design

Figure 3.1 illustrates a screen which implements the proposed colour interface design. Figure 3.2 shows a similar screen, with the same number of colours, that does not follow this design. In the latter, colour inconsistency is deliberately forced for demonstration purposes. Although the use of colour is not random in Figure 3.2, it is not consistent across applications running in different windows or between all icons and menus.

Obviously it is difficult to know what is the appropriate comparison to use against the design illustrated in Figure 3.1. Notwithstanding this difficulty, Figure 3.1 illustrates

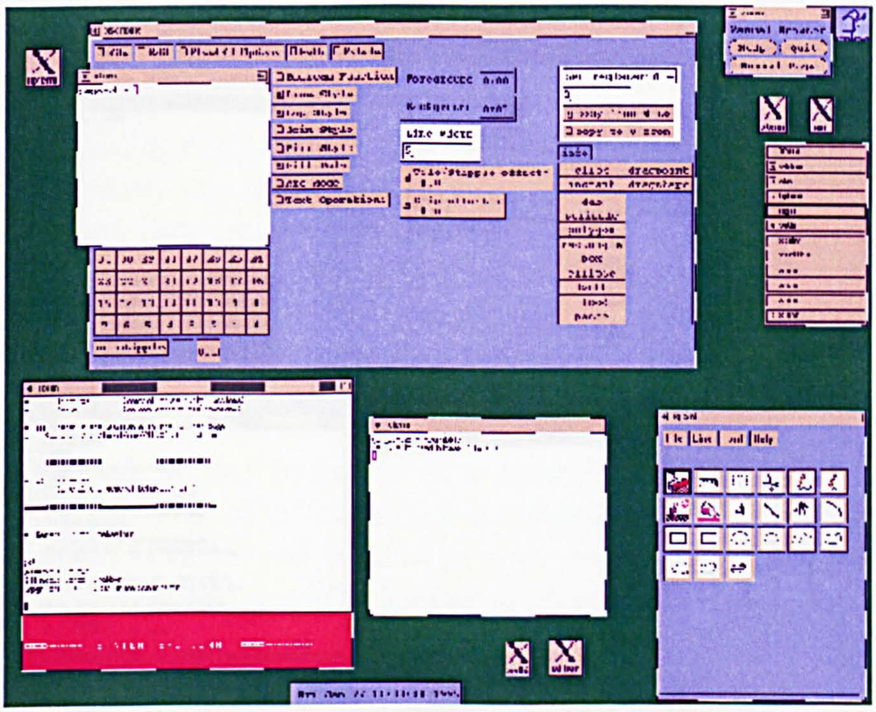


Figure 3.1: An implementation of the colour design on a UNIX workstation. The plate gives a user, *at a glance*, consistent information concerning screen areas that accept mouse input (yellow), text input (white), no input (blue), areas with hidden options (green) and urgent information (red).

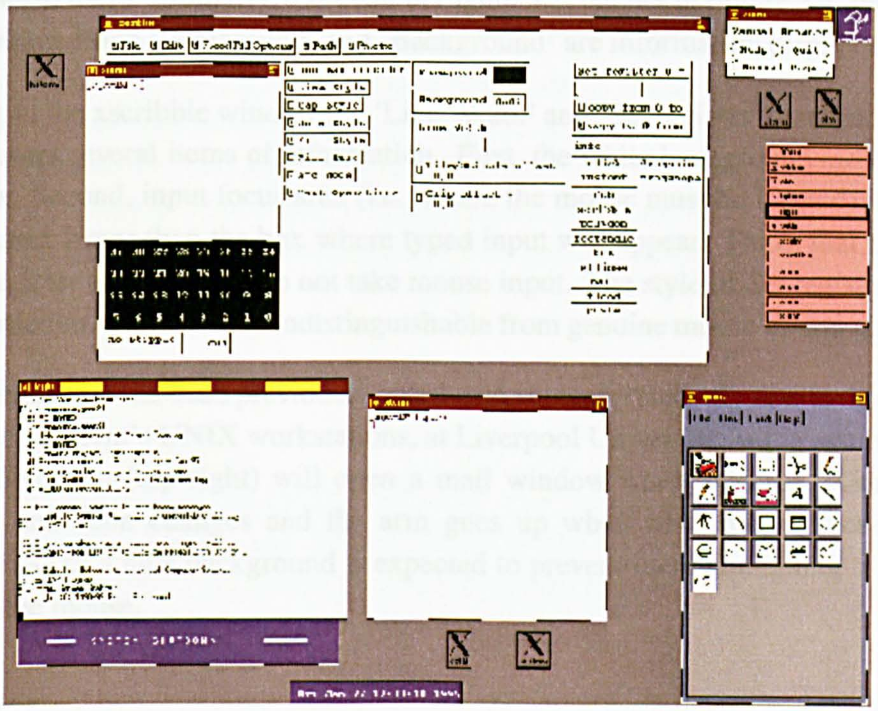


Figure 3.2: A colour plate on the same UNIX workstation without consistency across multiple application windows, prior to colour design implementation.

some immediate advantages of consistent interface design. For example, the 'xscribble' software interface (see the upper, large window) immediately allows users who know the colour design conventions to distinguish mouse selectable areas (yellow or green) from text input areas (white) and from areas that take no user input (blue). In this window, the colour design usefully differentiates titles (i.e. information) from the yellow mouse menu options even though the xscribble's usual user interface does not make this distinction perceivable to users. Similarly, it is clear in Figure 3.1, but not in Figure 3.2, that the areas of the window titled 'Foreground' and 'Background' are information display areas only.

Similarly, in the xscribble window the 'Line Width' and 'Set register' text fields in Figure 3.1 give users several items of information. First, the white background indicates a text input area. Second, input focus area (i.e. where the mouse must be located) is shown as being an area larger than the box where typed input will appear. Third, that Line Width and Set register are titles and do not take mouse input. The style of Set register in Figure 3.2, in particular, appears to be indistinguishable from genuine mouse option screen areas.

One user error that has been previously noted with students relatively new to the Computer Science department's UNIX workstations, at Liverpool University, is the assumption that the mail-box icon (top right) will open a mail window when selected. Generated by 'xbiff++', this icon changes and the arm goes up when new mail is received. The consistent use of a blue background is expected to prevent users attempting to select this icon with the mouse.

3.2.5 Implementation Issues

It is possible on a UNIX workstation running X11 windows to set individual widget background colours in applications, such as xscribble, by using tools such as "Editres" or by editing source code. While this allows much of the above design to be implemented so that all software running on the workstation will obey the conventions as far as possible, the convention for red, in particular, cannot be implemented in this simple way. There are also some difficulties in changing colours dynamically as described in Chapter 2. Software for common computer systems that converts existing coloured software to the high level design convention would obviously be a desirable approach to establishing the conventions.

This colour design could be considered as the initial or pilot product of the research. The overall aim is to propose a colour scheme for the design of the Web pages, which is the main product of the research. Before a high-level colour design for the Web (see Chapter 4) is finalised it was felt necessary to conduct empirical research, for example testing the effect of current bindings on user performance, in order to accommodate and reconcile current design guidelines, and other relevant knowledge as far as possible. It is likely, however, that the power of convention is far greater than minor psychological effects. The proposed colour design therefore was subjected to empirical evaluation, as described in Chapters 6, 7 and 8, because once released and established, such conventions become highly viscous i.e. difficult to change [Green 1990]. The empirical research reports several interesting results. For example, in subjective evaluation of the proposed design(s) users agreed that in most situations consistent design is desirable. This research is concerned only to explore novel functional uses of colour and its effect on user performance (i.e. in what manner and to what degree). Other related questions, such as that of specifying alternative optimal colour sets for target areas fall outside the scope of this research.

3.3 Summary

Consistency has received the attention of many user interface designers and researchers for over a decade. Baekar and Buxton [Baecker and Buxton 1987] provide a logical answer for wanting more consistent interface styles since they: "... accelerate the process whereby novices begin to perform like experts."

In this chapter, a simple scheme has been proposed which associates certain functions to certain colours. The scheme sets the mapping as categorical between the target feature and the target function. The implementation of the scheme was driven by consistency, which facilitates easy recognition (i.e. type of input) by associating the same colour with the same information and therefore results in the same user action.

Chapter 4

Novel Functionality of Colour in Web Pages

4.1 Foreword

"The Web requires designers to think about colour in new ways" [Niederst 1999].

This chapter describes the main product of the research - "A colour scheme for designing Web pages". Having developed a pilot scheme that uses colour to communicate different types of semantic information on the traditional UI platform (section 3.2.2), the research aims at delivering another colour scheme(referred as the main colour scheme) in order to investigate further uses of colour 'in a semantically complex way' (see Thesis Objective, section 1.1). The complexity of mappings varies between these two colour schemes. In the pilot scheme, the colour mapping could be described as 'simple'. That is, a single function (input type) is mapped onto the perceptual feature (colour). In the main colour scheme described later in this chapter, the colour mapping could be described as 'complex'. That is, a complex semantic function is mapped onto the perceptual feature - colour. The design, delivery and evaluation aspects of the main colour scheme are based on the pilot scheme and the empirical studies presented in Chapters 6, 7 and 8. The Web platform was chosen because it provides an ideal environment for testing semantically complex associations.

The issues concerning UID (User Interface Design) and WPD (Web Page Design) have been reviewed in conferences, journals and several text books [Nielsen 2000, Lynch and Jaffe 1990, Dillon 1991, Edwards and Hardman 1989, Gygi 1990, Landow 1989]. There are some fundamental differences between the designers' model(s) adopted in traditional UI design and the Web. The former is typically a front end designed for a software application. The software vendors create a user interface designed by either the software engineers themselves or specialist designers. It is generally the case that a small group of specialists develop UI products, which are then delivered to a large group of users. At the most, these users need to learn only the designers' model[1]. The Web, on the other hand, is an application linking various documents on the Internet. A HTML document, for example, serves as an interface for accessing various resources. A Web author creates personal HTML pages and accesses the content designed by others. A Web author is therefore a product developer as well as an end user. The unidirectional nature of a vendor exclusively delivering a product and an end user exclusively using it no longer exists on the Web platform. To date, many thousands of HTML authors are involved in designing Web pages both for personal and commercial use. This increasing number of designers has resulted in many designer models and many authors do not even have a consistent model. They design with a view for what looks impressive for a casual user rather than for ease of regular use. Web users must cope with this plurality of models.

The core features of the Web colour scheme described in this chapter are - novelty, functionality, simplicity and consistency. The delivery of the colour scheme was mainly driven by the last two features. The product is straightforward in that it uses only three colour rules consistently across all Web pages. The design aims at delivering a product that can be easily implemented by any Web author. Initially, the chapter discusses (sections 4.2 to 4.4) some general concepts and design issues in creating Web pages and attempts to identify problems requiring solution. Following this, a plausible colour scheme is proposed and discussed with respect to some problems frequently reported by Web users. The validity of the product is tested in a series of empirical studies in Chapters 9, 10, 11 and 12.

4.2 Problem Definition

The perceived benefits and opportunities offered by multimedia are bringing the Web to a different and much larger audience than hitherto [URL 1998, URL 1998]. According to the World Wide Web Consortium (W3C), "energetic efforts" are being made towards bringing the Web to a truly world wide audience in terms of supporting the display of all the languages of the world, with all their unique alphabets and symbols, directionality, and specialised punctuation [Niederst 1999].

[1]Frequent use of a limited number of systems allow transfer of knowledge and consequently, provide user training. For example, the suites of tools available in Microsoft office share common interface features, which are helpful to users.

Availability of latest technologies and techniques for adding enhanced interactivity to Web pages has taken Web design to new heights. For example, plug-in technologies such as the Flash and Shockwave players from Macromedia, Java applets, Scripting languages such as JavaScript and VBScript, Server-side application development technologies such as Java Server Pages (JSP), Active Server Pages (ASP), Perl and Coldfusion. Other alternatives for achieving control over document presentation and adding interactivity include Cascading Style Sheets (CSS), Embedded Font Technology, Dynamic HTML (DHTML) and Extensible Markup Language (XML).

Lynch [Lynch 1994a] captures the essence of Web design saying, "The purpose of Web design is to provide screen displays that create an operating environment for the user, forming an explicit visual and functional context for the computer user's actions." Rich interactive experiences for users are created with text, graphics, sound, and/or video images. Unlike traditional, stand-alone interface design where the screens are fixed, and users can be counted on to follow only a few relatively prescribed paths, the non-linearity [Waters and Mundy 1996] associated with Web documents and the costs associated with transmitting information from servers, creates unique opportunities and challenges for designers. For example, weighing trade-offs about how to create links between different kinds of information, what categories should lead to what other categories, types of navigational support that can prevent users from getting lost or disoriented, and how to balance the relative costs of using resources that require long transmission times, file space, and/or external helper programs to view them. It is necessary for the Web designer community to focus on pragmatic solutions to the problems in navigating through an incessantly increasing number of Web documents.

Visual Cues: The Web designer community has been investigating the need to provide visual feedback as an integral part of adding interactivity to Web pages. Client-side scripting languages such as JavaScript provide facilities to add interactive buttons, also called "rollovers", that change when the cursor or pointer is 'positioned over' them. By making a button light up (i.e. calling attention to itself), it is possible to provide a strong visual signal indicating whether the images or image areas are selectable or not. Rollovers can also be used to pop up a graphic with additional information about the link. Another form of providing feedback is through "status line messages". When the user places the cursor or pointer over a link, a message can be displayed in the status bar - the bar at the bottom of a Web browser such as Netscape that shows URLs (Universal Resource Locator) or displays the message "Document: Done". Other current practices include underlining hyper links in different colours and changing cursor shape when placed on a link area. Other than colouring text hyperlinks, all other practices require the user to position the cursor or pointer over the different parts of Web page to obtain some feedback. That is, users will need to spend unnecessary time in investigating the type of contents and links in the Web page.

At a Glance Visual Support: In general, electronic hyper-documents appear to provide less visual support and lack some of the physical cues that users take for granted when assessing paper based information. Furthermore, the physical appearance of a Web page is not fully under the control of the designer, indeed, even the read permissions of individual Web pages are differently restricted across users. It is often the case that a Web page provides relatively few cues as to : what can be selected, where a selection will lead to, how much information is at the other end of a link, what kind of information will be downloaded from the remote site, how long the downloading will take and how the linked information relates to the current page. The last provides a classic illustration, after Norman [Norman 1990a], of the potential for mismatch between a 'Designer's User Model' and 'Users' model' as what is presumably a rational link within text for the designer is sometimes a surprise when traversed by an end user. This is in contrast to the traditional UI in which learning the designers' user models allows users to come to know what to expect. At present, one can reasonably argue that there is no straightforward, widely accepted and adopted mechanism to get the above types of information from a Web page without traversing the link. Even the visual cues, for example rollover techniques and changing cursor shape when placed on a link area, do not offer necessary information to judge a resource to be accessed *at a glance*, i.e. without further investigation. Consequently, in the absence of such mechanisms, navigating the Web can place heavy cognitive loads on users [Dillon 1991].

Current user interface standards, style guides and guidelines have evolved over the years and were often written with particular application software in mind. The standards committee, such as the W3 Consortium (www.w3.org) has been making sincere attempts to incorporate guidelines for the integration of text, graphics, hypermedia links [Lynch and Jaffe 1990], and audio-visual media within computer software documents. At present, over a hundred guidelines and style guides are published for Web page design [Lynch and Horton 1997]. A careful examination these guidelines indicates that further research is required in several design areas, for example providing *at a glance* support in presenting different types of information.

Given the importance of the problem of informing users of different types of information, in advance, without selecting a link and moving the cursor or pointer around the Web page, it is desirable to deliver products (e.g. methods, style guides, guidelines) that would provide as much *visual support* and context as possible to help users, and to minimise the disruptive effects of sub-optimal network traversal.

4.2.1 Who is Doing Web Design? Pragmatic Solutions for Web Page Designers

Given that a large number of Web pages are created by Web authors with limited knowledge of online screen design issues, it is necessary to develop simple products (i.e. design solutions) that ensure functionality and consistency in designing Web pages. These products need to be pragmatic, inexpensive and easy to understand and implement in order to receive the attention of the Web designer community. The research has attempted to design such a product using colour, after an extensive literature review of this visual feature (Chapter 2).

The following sections provide analysis of some Web design issues, the design and delivery of simple Web products with respect to some common problems reported by Web users. A plausible, functional three colour scheme was proposed to facilitate easy identification of different resources (i.e. information), *at a glance*, while browsing Web pages.

4.3 Problem Analysis: Web Page Components

In general, a Web page includes a number of different multimedia components, for example: text, graphics, audio, video and hyperlinks. While text and graphics are supposed to convey information, hyperlinks are mainly used for Web traversal. Hyperlinks are simply pointers in one document to elsewhere in the same document or to another document, perhaps on a totally different server, which might be anywhere in the world. Hyperlinks are generally distinguished from the non-hyperlink components by: a piece of underlined text; graphics such as buttons, bullets, images (animated, static, dynamic) and icons; and colours or reverse video.

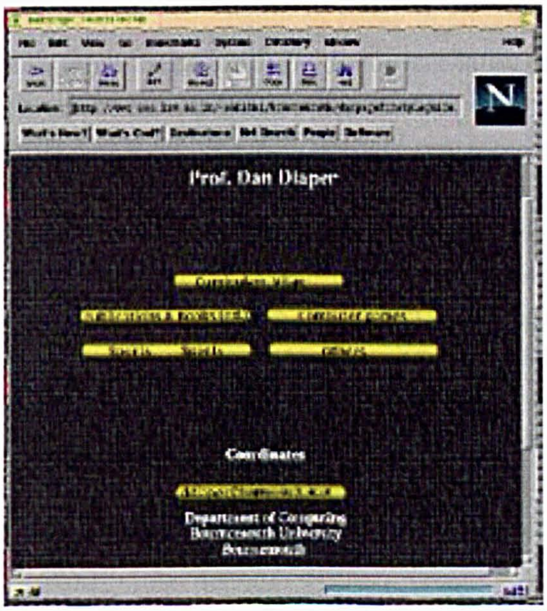


Figure 4.1: A screen dump of a Web page that does not follow the three colour scheme. The display includes six buttons out of which two are not selectable. The design does not provide any visual cue concerning screen areas that are not selectable by a user, *at a glance*.

4.3.1 Resources on the Web

Hyperlinks point to different resources on the Web. A resource is a document containing information in the form of text, graphics, audio or video. Resources vary in content and size and are downloaded whenever the user activates hyperlinks that comprise locations of URLs with a keystroke or a mouse click.

In a sea of Web documents, it is impossible to know in advance what a document is until a request is made and in most cases until the document is loaded. This means that users are unaware of the content of the resource when making a request. They normally look for the relevant keywords and select the link without any idea regarding the downloading time and content of the resource that the link points to. This may often cause several problems for the users, for example, users may end up with dead-end pages with no options, under-construction pages, pages with heavy and unwanted data dumps, pages with restricted access, large graphic, audio and video files, pages with a complex knit of menus that demand users' time to delve deeper into the site which may be highly irrelevant for many of the users. Consequently a great deal of their time can be wasted in following such links during Web browsing. Search engines provide an excellent example, in that it is often hard to choose which of the links to follow up.

4.3.2 Graphics

Graphics images are an integral part of the user's experience with most Web sites as it is widely believed that images can add a lot to the visual appeal as well as to the information content of a Web document and their complete absence is noticeable to most, non-neophyte Web users. Well-used graphics can crystallise a presentation for a reader, providing a critical catalyst for understanding [Lynch 1994b]. Used poorly, on the other hand, images can confuse the audience, distract from or even render mute a message. Overuse of graphics leads to a "Clown's pants" effect where everything is garish and nothing is really emphasised.

In a Web page, it is possible to represent links using graphics. That is, placing a graphic within HTML anchor tags will make the entire image a link to a single document, regardless of where the user moves the cursor or pointer on the image. In a Web survey conducted by the author, a sample of 350 Web documents (see section 4.4 for details) were examined. It has been found that approximately 55% of the hyperlinks are tied to graphics of some sort, rather than to the HTML standard of an underlined portion of text. One reason could be that designers believe that text will not compete for the viewer's visual attention in the same way as graphics.

A major problem with graphic hyperlinks is that users often find it difficult to identify the graphics that are *selectable* from *non-selectable* ones. The example Web page[2] in Figure 4.1 includes six buttons out of which only four are selectable. All the buttons look alike and there is no immediate visual support for the user in identifying selectable ones *at a glance*. Obviously, the greater the density of text, graphics and hyperlinks in a Web document the more difficult a selection becomes for users. The problem is further intensified with the use of image maps.

4.3.3 Imagemaps

It is also possible to create multiple links (also called hot spots) within a single graphic. These graphics are called imagemaps[3]. The effect is created using HTML tags or scripts on the server.

In today's Web browser implementations, imagemaps (i.e. the selectable graphics used on Web pages) often present a rather poor user interface unless the image itself has been crafted so as to have well-delineated 'active' regions. Often there is no clear indication of where a user should click. The imagemap from the Web in Figure 4.2 provides an example of this problem[4]. Apart from the fact that it is quite slow to load because of its size, once loaded there is no indication to the user that it is anything other than an image. Actually, the bottom right hand corner contains a small, selectable button around the region containing the painter's signature. It is very likely that many users will not realise that there is a hyperlink button within the imagemap, and that this will be reinforced if they click anywhere on the image other than in the very small active area. The design in Figure 4.2 is particularly odd since with only a single button there seems no good reason why the selectable area should not be coextensive with the full imagemap's extent. It

[2]This page has been deliberately created for the experimental studies described later.

[3]These imagemaps are of two types: client-side and server-side. In the first, the processing takes place in the browser on the user's machine, thus client-side. In the second, the map information resides on the Web server and is processed by the server or a script. The main advantage of client-side imagemaps is that they are self-contained and do not rely on the server. As a result they cut down load on the server and improve response times.

[4]This picture is no more available on the Web in the same state described here.

seems fairly obvious that users should be informed about the selectable regions as quickly as possible and that colour, not least because it is an under used interface feature, is a possible candidate for such a role.

Figure 4.2: A scaled down screen dump of a large image (15 MB) on the Web. The display does not provide any cue concerning screen area(s) that are selectable by users. Besides high downloading time, it takes a considerable amount of time for an average user to identify the selectable area in this image map.

One of the important problems with pictures is that they take a long time to load compared to text. In the sample of Web pages mentioned earlier (see sections 4.3.2 and 4.4) about half the documents with graphics have pictures carrying over 50,000 pixels, many of which take up thousands of kilobytes. People who design such pages presumably intend to entice users by using captivating graphics effects. While surfing the Web, users will often find what they perceive as non-functional, albeit elegant, Web documents, for example, Figure 4.3 from: *http://utwwu2.wb.utwente.nl/wouter/dots.html* (as appeared on 10 Sep, 1998). The Web page contained only the dots and no text or hyperlinks.

In the long run, people may not be impressed by such non-functional graphics effects if they must spend a considerable cost and time downloading them. The concept of non-functionality is, of course, difficult in that what may be non-functional to some users may not be the same for others, depending on users' goals, style, and so forth. What is undoubtedly vital is that it is the users' perception, not the designers', that is important. During informal observations in undergraduate computer laboratories, there have been

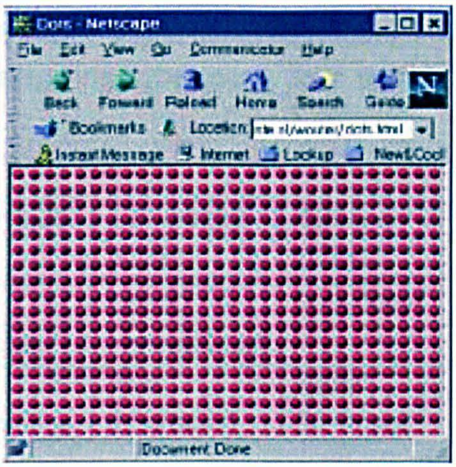


Figure 4.3: Non-functional image dump

several instances of negative emotional responses from users after waiting, sometimes minutes, to receive a Web document that they ultimately perceive to be primarily, or even exclusively, consisting of gratuitous graphics.

At the least, designers should seriously consider the effect of where they locate graphics in Web documents on those who may be using the Web with less than ideal platforms (e.g. any access to the Web via a modem). It is highly desirable that users are warned if any hyperlink leads to a large graphic or other slow loading object. In the recent times, designers have been adopting different practices such as (1) using a thumb nail graphics as a representation of the original large graphic. This practice provides a warning of image size prior to downloading and (2) a small text message describing the size of graphics is displayed in the vicinity of a link. These practices indicate the increased awareness of high downloading times associated with graphics although it is still the case that many Web authors are yet to adopt them seriously in designing Web pages. The same argument can be extended to using appropriate graphic formats (e.g. GIF, JPEG) in creating and storing graphics for the world- wide audience. Ideally, Web designs should give users the choice of a sans- graphic mode. Such a mode might sometimes suit even those users with state-of- the-art platforms and Internet access when access time is of vital importance. Disabling graphics locally in the browser used to be an option, but now many sites can not be understood or navigated without pictures.

In summary, little immediate help is provided to indicate the content and/or accessibility of resources in advance. A better organisation of resources would greatly reduce cognitive load associated with navigating cyberspace, save users' time and hence may increase users' satisfaction.

Resource attribute/ criteria	*Category 3* (C3)	*Category 2* (C2)	*Category 1* (C1)
Location	Local	Non-local	-
Read permission	Unrestricted	-	Restricted
Load time (average times)	Fast 0-30 sec.	Slow 30-90 sec.	Very slow More than 90 sec.

Table 4.1: Accessibility of a resource is defined in terms of the three attributes listed in the first column. Each attribute can be graded on any or all of the categories C1, C2 and C3.

4.4 Product Design: Resource Classification

Based on the previous analysis and the author's informal interviews with several people regarding their experiences on the Web, a plausible list of features that would affect either (a) the *accessibility*, or (b) the likely *desirability* of information available on the Web was constructed. These features were used in the analysis and classification of the Web pages. The list of features (see sections 4.4.1 and 4.4.2) is not necessarily exhaustive. Tables 4.1 and 4.2 show all the features considered in the present analysis. There are several other important features that were not considered in sufficient detail in the current analysis, for example, delays associated with underlying network architecture and performance of servers. Detailed analysis and inclusion of these features would be useful; however, the thesis aims at delivering a simple product described in terms of a small set of features.

Having identified the features, a Web survey was conducted in order to determine a representative sample of Web pages and analyse them with respect to each of the features. Infoseek, Yahoo and AltaVista search engines were used with the search terms: "colour", "color", "graphics" and "user interface design" which are relevant to the current research. The search engines produced hundreds of hits. Since these documents were expected to be

Resource type	*C3 (high desirability)*	*C2 (medium desirability)*	*C1 (low or undesirable)*
Text	Text maximum of 4, screens (users prefer to read only small amounts of text)	Text over 4 screens	Data dumps, Foreign language pages
Graphics	with small size graphics (less than (10,000 pixels)	with medium size selectable graphics (10,000-50,000 pixels)	medium size non-selectable graphics or large graphics (over 50,000 pixels)
Audio/Video /animation		medium desirability due to network load/ plug-in requirements etc.	
Interactive features (CGI / Forms etc.)		users may be less inclined to fill up data	
Miscellaneous		Redirection links, Menu of links File download	Under construction pages

Table 4.2: Classification of Web resources according to desirability criteria.

biased towards graphics content, another random category of search was carried out using the term "poetry" with an intention of obtaining more text oriented documents. While no strong claims can be made that the resulting documents cover every aspect of the listed features, the first 350 documents have been selected. Of these, 150 were from the first category and the remaining from the second. They formed a sample of current Web page design practice.

The Web pages varied from text-only documents to graphics dominated pages. Some documents filled a single screen while some others were several times larger. Several documents had very small graphics regions (less than 100x100 pixels), many of them being selectable icons. Many others had larger graphics (over 200x200 pixels) to add colour and visual appeal to the pages. The geographical locations of the resources also varied widely, from UK (the author's address) to as far away as Japan. The pages were examined with respect to the list of features mentioned earlier.

Out of 350 documents, 191 contained graphics images along with some text (around 55% of the total); 122 were text-only documents (35%); 30 dead or stale links, 5 redirection links and 2 under construction pages. Among the 191 graphics documents, 76 contained very large images, 52 had medium sized graphics and the remaining small graphics. Among the text-alone documents, 51 were small (less than 4 screens) and the rest were very large. In this survey no audio/video resources were found[5].

As can be seen from the above statistics, the graphics usage was widespread, even in the "poetry" pages, which were predicted to have a relatively low graphics content. The images came in various sizes affecting the downloading times. Further, the text-alone pages also came in various sizes. There were several dead-end pages and data dumps. The survey indicated that there was little visual support for the users to assess the type of information they access. Few visual cues are provided to indicate any of the features considered in the survey which were expected to influence either the accessibility or the desirability of the information.

Based on the above survey and the features discussed so far, this section proposes a plausible classification of Web documents into several categories. Each of these categories (see Tables 4.1 and 4.2) indicate the accessibility and/or desirability of content of the linked-to resource and allows people to decide whether they wish to spend time downloading the

[5]Since the survey in 1997, there has been increase in the use of audio/video on the Web because of the popularity of technologies such as Macromedia. The vector graphics model adapted by Flash delivers high quality pictures much more compactly than pixel based pictures. The audio segments in Flash files, however, still significantly contribute to file size.

document, without moving the cursor. This is especially helpful in situations where an expectation is created that graphics in a series are of a certain size, and then that expectation is betrayed when the image turns out to be much larger. The final product is a three colour scheme and is described in section 4.5. The colour scheme represents the categories in order to provide a simple mechanism to indicate to users the nature of resources, *at a glance.*

The analysis presented in the following subsections illustrate how the final product is designed. It should be noted that users are not required to learn the inner details of the analysis in order to use the final product.

4.4.1 Accessibility

Accessibility defines ease of access of a resource from a Web page. Some links have unrestricted access to all users world wide, while others - for example, manual pages for setting up a local company Web site - have restricted access. The thesis attempts to describe accessibility in terms of three features - Location, Read permissions and Load time, for the reasons discussed below. Each of these three features in turn depends on several other factors. Read permissions of a resource indicate access restrictions. The delays associated with accessing a resource are encapsulated in location and load time features.

4.4.1.1 Location: In general, when a Web page is accessed, the client's perceived response time has two components. One of the components is network delays. The other component is queuing delays at the server, covered in section 4.4.1.3.

Network delay is the time for application data to be transmitted between the sending and receiving machines. Contributors to the network delay include the network bandwidth and delays caused by distance or congestion [URL 2000a]. In other words, the speed with which one can access a resource may partly depend on its location. For instance, resources located on a local machine (i.e. server and client machines) could be more readily accessed than those on machines located externally in the sense that network delays will be lower in the former. Estimating network delays requires a priori knowledge of the underlying network architecture - the topology of network, the routers connected to the network, the type of routing protocol - its complexity, demands on memory and computation, the limitations of the protocols such as the maximum number of hops, link costs etc. For the purpose of illustration, in networks that use Routing Information Protocol (RIP), routing decisions are based on number of hop counts (i.e. number of routers). In this, the path with the lowest hop count is always preferred even if the longer path has a better aggregate bandwidth and slower delays. Under these circumstances, considering a hop count as a measure in estimating speed of accessibility would become a debatable factor. On the other hand, networks using Open Shortest Path First (OSPF) protocol, the routing domain is treated differently and there is a possibility of using non-optimal routes, which may introduce additional network delays. In case of the location feature, with bandwidth being the same or constant in a local site, it would be reasonable to assume that the resource to be accessed may not be too far away. In other words, hop count[6](i.e. number of routers) will be low. Estimating speed of accessibility based on location features may therefore be considered an appropriate feature for the current analysis.

4.4.1.2 Read Permissions: Some organisation-specific Web pages or sites that provide valuable resources to customers, ask for a valid username and password before granting read permissions. For example, university library catalogue information at *www.unl.ac.uk*, technical information at *ligwww.epfl.ch* and journal papers and conference proceedings at *www.siggraph.com*. Among the restricted category, there are Web sites that never allow outsiders to access a specified set of resources, for example *ligwww.epfl.ch* and *www.nasa.gov*. The other type, for example journal papers and conference proceedings sections of www.siggraph.com provide facilities for the interested users to register with them instantly to enable them to access the information. It makes some sense to group such sites under a separate category while determining the degree of accessibility. One point to note is that people would not know about the restricted access until they browse through the link(s). Any scheme that uses this feature in evaluating the nature of Web resources would help a Web author in that he/she could decide to use the scheme to indicate such links i.e. with restricted read permissions in their Web pages.

4.4.1.3 Downloading Time: This is another feature that influences accessibility. For the purposes of categorisation, download time is taken to be the time taken between user requesting a resource and response seen by the user, assuming resources are located at the same location. Location dependent aspects were considered in section 4.4.1.1. This research attempts to analyse downloading time in terms of two factors: resource type and features that influence server load. There are other related factors such as network connection[7], which are not considered in the present analysis.

[6]The issue of maintaining multiple mirror sites at different geographic locations comes into picture partly due to minimising hop count. Consequently, users are advised to choose a nearby Web site.

[7]It is difficult to accurately determine the speed of a network connection at any time. Also, different users accessing the same Web page may experience different network delays. Crovella and Carter [Crovella

Type of Resource:

A request for a certain type of resource[8] can influence the download time. For example graphics (see section 4.3.2), audio and video resources are usually large compared to text documents. With streaming audio and streaming video technologies, the audio and video can begin playing while the files are downloading. While streaming technology improves downloading time, these resources are generally time consuming even on local sites and hence can be considered less readily accessible than simple text resources.

Features that Influence Server Load: The other component that affects the downloading time is queuing delay at the server. While size of a resource is a significant factor in download time, there are server dependent delays which do not depend on file size. Server-induced delays can be difficult to uncover and quantify in simple as well as complex environments. The delay problems may have a number of sources, such as slow Domain Name Server (DNS) lookups, slow partner servers (i.e. ad servers), the Web server (overloaded due to processing too many small graphic i.e. gif files, use of frames or Server Side Includes) and a back-end database server [URL 2000a]. Some features in the Web page design could make a significant difference to the server load. For example, Web experts have indicated [Deitel *et al.* 2000, URL 2000a] that use of frames, Server Side Includes increase the load on the server. A large number of frames on a page may significantly increase the server load in that four requests for four 1K files (the frameset and the contents of three frames) would mean more work for the server than a single request for a 4K document [Niederst 1999]. Similarly, CGI[9] (Common Gateway Interface) processes, which are very resource intensive. These features influence server load and consequently the download times, irrespective of the actual resource size.

2000, Carter and Crovella 2000] have established the need to know bandwidth characteristics along a path for efficient prediction of transfer time or response time. Several tools such as pchar, putchar, clink were developed to accomplish this task. A comparison of the tools, for example, pathchar and clink indicated that for some links, the bandwidth estimated by them differ by as much as 200% and each of them took 20 minutes to run. That is, it would take a long time for a client to estimate the bandwidth characteristics and the result may not always provide a reliable estimate. While difficulties such as these make it very difficult to consider this feature for analysis in determining the accessibility, some indicators such as type of resource could be taken as guiding factors in assessing the downloading time and consequently the accessibility.

[8]The resource dependant delays exhibit themselves in relationship to one or both of the physical performance measurements, namely network and server delays.

[9]Use of ISAPI (Microsoft Internet Server Application Programming Interface) programs do not have the overhead of CGI (Common Gateway Interface) applications since they are loaded into the memory at the server start-up. Similarly, a Java servlet, which is considered a 'light process', reduces server load compared to a CGI scripts performing the same task. With Dynamic HTML, many visual effects can be implemented directly in the client-side browser thereby reducing server-side processing delays.

The reason behind performing this research exercise is to understand different types of delays that might affect the accessibility of Web pages and provide some helpful guidelines about evaluating and classifying different types of Web pages. From an average Web author standpoint, this help should enable him/her to relate resource dependant delays with server load and make a decision on downloading time. For instance, the download time will be high when a request is made for a video file, a page with multiple frames or a CGI script. The thesis attempts to provide such help by performing an analysis of the factors that affect the downloading time.

There are many other features that the Web experts would consider to be critical in determining accessibility. For example, the actual speed of the machine, the type of browser, the type of display, using proxy servers. In general, the speed of a user's machine has little bearing on downloading time unless there is not enough RAM or there is constant swapping to disk. As far as the proxy servers are concerned, they sit in between a client application, such as a Web browser, and a real server. They intercept all requests to the real server to verify whether they can fulfil the requests themselves or forward the request to the real server. One main advantage of proxy servers is to improve network's performance by functioning as a caching server. Using cached Web pages, the proxy servers will serve already-accessed Web pages to requesting clients without requiring outside access to the Internet thereby saving considerable amount of network cost and connection time [URL 2000b]. Inclusion of all these features will elaborate and complicate the process of determining the accessibility. Consequently, these features are not fully explored in this thesis as the research aims at reducing the complexity of determining accessibility by considering a small set of features.

Based on the three features discussed so far - that is location, read permissions and downloading time - an attempt has been made to grade the accessibility of a particular resource into different categories. Table 4.1 shows three plausible categories C3, C2 and C1 representing high, medium and low or no accessibility respectively. It should be noted that the scheme of categorisation provided in this section provides only approximate estimate.

A single resource link can be graded on any or all of the attributes listed in Table 4.1. For example, a link can point to an audio/video resource on a local site. In this case it has high accessibility (category C3) by the 'location' criterion but poor accessibility (category C1) with respect to the 'download time'. A plain HTML resource may be placed under high accessibility. However, same sized HTML resource with multiple frames increases perceived download time and may be placed under low accessibility.

The research aims at obtaining a single category representing overall accessibility of a resource. It is therefore necessary that various grades, C3, C2 and C1, of all resource attributes need to be combined to produce a representative value. A resource can be classified according to a tree structure branched into three levels (Figure 4.4). The first level bifurcates at the root according to the location criteria of C3 (local) and C2 (non-local). The second level again produces a bifurcation at each node for read permissions - C3 (unrestricted) and C1 (restricted). At the third level, each node branches into three categories representing average load times - C3 (0-30 sec), C2 (30- 90 sec) and C1 (more than 90 sec)[10]. C1, C2 and C3 are assigned values 1, 2 and 3 respectively for each attribute contributing to accessibility. Higher score indicates higher accessibility. Each branch in the tree from the root to a leaf node represents a resource categorisation on all three attributes. A final category value can be obtained for each branch by merging all the three category values.

A simple score based approach is designed to distribute resources into three final categories C1, C2 and C3. This scheme adds up the values from each of the attributes (i.e. location, read permissions, load time) and assigns a category to the final score. The maximum score that can be produced by this scheme is obtained from the left most branch with all C3 values: 3+3+3 = 9. The minimum value is obtained from the right most branch C2-C1-C1: 2+1+1 = 4. The possible final scores therefore range from 4 to 9 to which the following final category coding may be assigned:

4-5 (C1), 6-7 (C2), 8-9 (C3)

According to this, the shares of C3, C2 and C1 categories are 3/12, 6/12 and 3/12 respectively. As can be seen, the probabilities are evenly distributed across the link categories.

While this scheme is assumed to provide a reasonable coding of various resources, some resource combinations that may be included in the scheme can be subjected to criticism. For example, it may not make much sense to talk about the slow downloading pages (e.g. C1) with restricted read permissions (C1). In this case, the resource is correctly assigned

[10] The Web literature review indicates that the optimum waiting time for downloading a Web page is 8 - 10 seconds [Nielsen 2000, URL 2000a]. Considering the different types of modem connections, for research purposes, an average time of 30 sec has been considered as the maximum acceptable time that a user can tolerate. This limit is set based on the opinions of a representative sample of eight users who are well versed with Web technology. During an informal discussion, these users have discussed their experiences while downloading Web pages from home (slow connections) and work sites (fast connections), and reported that the mean time spent in downloading pages is more than 8 sec.

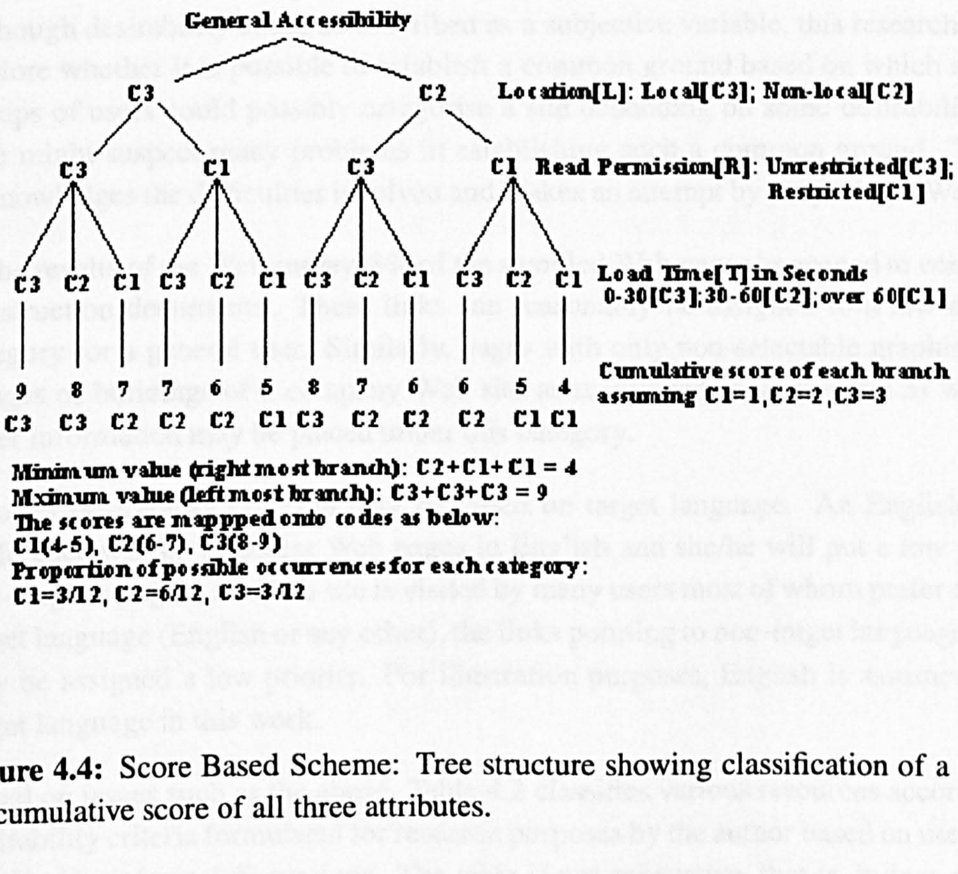


Figure 4.4: Score Based Scheme: Tree structure showing classification of a link based on cumulative score of all three attributes.

C1 indicating that the degree of accessibility is low, that is, the scheme attempts to make some sense by not saying that these pages are readily accessible. Combinations such as these are not fully explored in this research.

4.4.2 Desirability

Desirability of a resource specifies the nature of the content that may or may not be of interest to the user. It may be user specific reflecting individual needs and preferences. Although desirability could be described as a subjective variable, this research intends to explore whether it is possible to establish a common ground based on which all users or groups of users could possibly categorise a site depending on some desirability criteria. One might suspect many problems in establishing such a common ground. This thesis acknowledges the difficulties involved and makes an attempt by proposing a Web scheme.

In the results of the Web survey, 8% of the sampled Web pages happened to contain under construction documents. These links can reasonably be assigned to a low desirability category for a general user. Similarly, pages with only non-selectable graphics (such as images of buildings of a company Web site as mentioned in section 4.3.3) without any other information may be placed under this category.

Another desirability criterion may be based on target language. An English speaker's preference will be to access Web pages in English and she/he will put a low priority on non-English pages. If a Web site is visited by many users most of whom prefer a particular target language (English or any other), the links pointing to non-target language resources may be assigned a low priority. For illustration purposes, English is assumed to be the target language in this work.

Based on issues such as the above, Table 4.2 classifies various resources according to the desirability criteria formulated for research purposes by the author based on user feedback obtained in informal discussions. The table is not exhaustive, that is, it does not include issues such as browser-specific Web pages, links pointing to resources such as newsgroups, VRML pages, pages using flash etc. Further, issues relating Wireless Application Protocol (WAP) Technology and associated Wireless Macro Language (WML) Script pages, WebTV are not considered for the purpose of analysis in this thesis. The selected desirability criteria were not subjected to rigorous evaluation but evaluation of the Web scheme (implementation of which uses desirability criteria) carried out by a representative group of users (consequently the validity of the criteria) is reported in Experiment 7 (Chapter

12). If the use of the criterion is justified (see Experimental Chapters 9 to 11), it could be verified further with respect to target audience based on some user modelling.

Following the convention in the previous section, category C3 has a high desirability value of 3 followed by C2 (medium desirability; value 2) and C1 (low desirability; value 1). It should be noted that Table 4.2 is not to be taken as a universal scheme for evaluating desirability, but only as a guideline. For example, graphics/audio/video are generally omitted from the high desirability category. Many online news Web sites (e.g. www.cnn.com) often have news in text as well as video clips. The text-based news is usually self-sufficient and covers all aspects of a news story. The video clips serve mainly as an accompaniment for visual presentation. In such Web sites, video, being an appendage, can be assigned only medium desirability as per the above table. On the other hand, companies specialising in graphics/animation would like to keep graphics/video as a main ingredient of their Web sites. Under such circumstances the table needs to be modified. The main point is to codify desirability into three appropriate categories C1, C2 and C3 based on the target audience. Note that unlike attributes listed in Table 4.1 representing accessibility, all entries in the desirability table are mutually exclusive. They generate only one level of category values unlike the accessibility tree structure (Figure 4.4). This simplifies the process of evaluating the total outcome.

4.4.3 Total Outcome

In the context of delivering Web content, Niederst [Niederst 1999] says that "... There is no absolute rule here". With the knowledge of the existing Web design guidelines, Web researches generally attempt to explore several alternatives for successfully deploying resources to general public.

With this intention, the thesis intends to explore merging both accessibility and desirability of a resource in order to produce a single category that can then be used as a basis for assigning colour to Web pages. Different types of Web pages exist today. Some of them are readily accessible but their content is less desirable for an average user (e.g. statistical dumps). On the other hand, some pages have highly desirable content but the speed with which one can access them can be low (e.g. server-side applications). The thesis assumes that people would generally attempt to choose Web resources that are readily accessible and also desirable, and consequently the Web designer community would prefer to indicate any such resources to the user. This assumption provides one possible reason for combining both into a single category.

Desirability (right) / Accessibility (below)	1 (C1)	2 (C2)	3 (C3)
4 (C1)	5	6	7
5 (C1)	6	7	8
6 (C2)	7	8	9
7 (C2)	8	9	10
8 (C3)	9	10	11
9 (C3)	10	11	12

Table 4.3: Combined result of accessibility and desirability values. The first column shows six levels of the score based accessibility ranging from 4 to 9. The first row shows three levels of desirability, 1 to 3. This results in a total of 18 combinations.

Table 4.3 shows the values of accessibility and desirability combined to form a single score. The range of category measures for accessibility and desirability are shown in the first column (from 4 to 9; this range was chosen based on the final set of values presented in Figure 4.4) and the first row (from 1 to 3) respectively. This table therefore presents all the eighteen combinations of accessibility and desirability measures. In the total measure, the value of each combination is obtained by a simple addition of accessibility and desirability values.

The total outcome takes eight distinct values ranging from 5 to 12. The lowest value represents Web resources with the lowest accessibility and desirability. The highest value represents the most desirable as well as the most accessible Web resources. The values in between represent a gradation of these attributes. These values from low end to high end can be encapsulated into three overall categories: C1=5-7, C2=8-9, C3=10-12. These three ranges partition the eighteen combinations evenly with each category appearing 6 times.

4.5 Product Design: The Three Colour Scheme

Colour is a predominant visual feature on the Web. Understanding of colour can benefit greatly if a developer knows how to use it in both functional and technical manner[11].

From a functional standpoint, Chapter 2 describes the uses of colour in current computer systems and designing Web pages. In section 4.2, existing practices in indicating various types of Web resources were briefly mentioned. For example using icons to indicate internal and external sites, fun sites, child suitable sites and colour representing visited, non-visited, active links. As this research intends to explore the uses of colour on the Web, colour is used as the main cue to indicate different types of Web resources.

This section describes the final product, the high-level Web colour design, in which the three resource categories discussed in the previous section are visually encoded using colour as the salient feature. The purpose is to deliver the colour scheme to users to enable them to assess a resource based on its accessibility and desirability *at a glance*, i.e. without even moving the cursor. It should be noted that the combined effects of colour alongside with other browser (and also desktop) interface features such as layout are not fully explored in this thesis.

Table 4.4 shows the high-level colour design, which illustrates a novel functional allocation of colour in designing Web pages. The table defines an apparently simple mapping from the categories C3, C2 and C1 to the colours Yellow, Green and Blue respectively. Each category however is the result of a combination of accessibility and desirability measures as described in the previous section. The colour coding therefore specifies a mapping between the perceptual category i.e. colour and the function i.e. the type of a link. This is in contrast to the pilot UI scheme (see section 3.2) in which the mappings were simple between colour and the function (i.e. input type). In comparison with the pilot product, colour in the Web product could be described as a semantically overloaded feature.

The thesis acknowledges the difficulties in characterising links into one of the three

[11]From technical standpoint, one interesting issue arises when colours from a 24-bit colour space need to be displayed on a 8-bit display. Browsers do not rely on the computer's system palette and reduce and re-map colours to their own built-in palette consisting of 216 colours. This guarantees that colours (e.g. images) almost look the same on all 8-bit systems and therefore greatly assist Web developers in designing sites. It should be noted that if the browser is running on a 16-bit display slight colour shifting and dithering might occur depending on the browser and operating system. On a 24-bit display, this palette does not come into effect and all colours will be displayed accurately [Niederst 1999]. It is also worth noting that, as per 1999 statistics, only 11% of people use 256 colour displays [Nielsen 2000]

categories based on accessibility and desirability. The characterisation is not very sophisticated, nor does it need to be for the research purposes, although it would be possible to have a more complex system of mappings. It is an experimental scheme that will be used to try to see if useful results can ensure.

Accessibilty / Desirability outcome category	Colour code mapping
C3	Yellow
C2	Green
C1	Blue

Table 4.4: The Web colour scheme: Yellow, Green and Blue colours (background/border) were used to indicate the accessibility and desirability of contents of the linked-to resources on the Web. Red is reserved for emergency signalling.

4.5.1 How to Use the Colour Scheme

The purpose of applying the proposed Web product is to assign colour codes to hyperlinks depending on the type of resources they point to. This is done in two stages. The hyperlinked resource should be first evaluated with respect to accessibility and desirability criteria. In the second stage, colour coding is assigned to the links based on the Table 4.4. For convenient usage of the Web product Table 4.5 may be used to map directly from the accessibility, desirability measures to the colour.

The measures assigned by a user may be based on his or her understanding of the nature of the resource. For example an audio clip on a local site could score high on accessibility, say 8. If the clip is not perceived to be very important to visitors of the site, then desirability may be assigned a medium value, say 2. The mapping in Table 4.5 produces the colour code G (Green) for the resource. Some resource types require different treatment. For example, under construction pages or stale links on local sites score low desirability value of 1, but a high accessibility value of 9. The tabular entry assigns colour code G for such links. In order to avoid this, irrespective of accessibility, these links should be assigned colour code B (Blue) if they are to be placed on par with links of low accessibility and

Desirability (right) Accessibility (below)	1 (C1)	2 (C2)	3 (C3)
4 (C1)	B	B	B
5 (C1)	B	B	G
6 (C2)	B	G	G
7 (C2)	G	G	Y
8 (C3)	G	Y	Y
9 (C3)	Y	Y	Y

Table 4.5: Mapping between functional categories accessibility, desirability and the perceptual category colour. The table uses the three colour scheme.

low desirability[12].

Consistency and predictability are two important delivery requirements of the designed product, aiding users in identifying the origin and relationships of the Web pages, providing access to interface and page elements, and a graphic design scheme. Simplicity is another essential requirement. The product design was driven by this requirement to comprise of only a minimal number of colour rules. The three colours blue-green-yellow as background or border are used to indicate the degree and result of accessibility and desirability of the Web components (links, graphics), *at a glance*. As with the UI colour design, the degree of accessibility and desirability of content increases from blue to yellow in the VIBGYOR spectrum (see Chapter 3). Colours are chosen from the basic colour set and shades from the browser-safe palette [Weinman and Heavin 1997a].

4.5.2 Advantages of the Three Colour Scheme

The Web is a relatively new platform concerned with the presentation of information in a non-sequential fashion through a number of hyperlinks. This interactive nature of the Web opens up many exciting new possibilities, but demands a high degree of discipline by users (both designers and end users) while browsing through different types of available

[12]It may look more appropriate to use a scheme of minimum(accessibility, desirability) in these cases. However, uniform application of such a scheme across the links produces codes skewed towards a particular colour.

information (i.e. resources). Primarily, users' tasks involve a significant amount of search and identification while selecting a desirable link in a maze of hyperlinks representing highly inter-linked resources. Experiment 3 (see Chapter 8) has demonstrated the effect of colour grouping in search and identification tasks. The grouping had a beneficial effect on users' performances. Considering that Web browsing involves large amount of search and identification, the combined effect of the proposed novel use of colour and its grouping effect was expected to demonstrate a significant influence on users' performance. The validity of the combined effect of colour is tested in a series of experiments described later in the thesis (Chapters 6 to 12).

The implementation of the product applies the three colours across the links, either as background or border colours, without affecting the original graphic content, if any. The design ensures a consistent approach to navigation and allows readers to quickly adapt to the design, and predict the speed of access and desirability of content across the pages of a site. The style of navigation evolves as a natural consequence of consistent and appropriate handling of the three colour set.

The proposed product supposes that many Web users will be using systems that result in relatively high downloading times (due to low bandwidth, system set up etc.). The users will therefore prefer to avoid long waiting times and unwanted content. One of the main goals is to make the documents usable and valuable to the large number of users by giving them a sense of the site's organisation by making the logic and order of the site visually explicit. The number of colours is kept to three so that the scheme could be implemented even on low-end graphics systems.

The potential benefits of the Web colour design may include:

- A reduction in the overhead associated with navigating by making the hyperlinks visible. The visibility occurs, *at a glance*, without requiring the user to move the cursor, prior to making a request.
- Consequently, increased opportunities for users to juxtapose and compare various links pointing to different information. Users can be warned about undesirable (e.g. pages with restricted read permissions) and/or slow links, before Web traversal.
- Web page designers may be assisted in presenting the style, content and navigation structure of the hyperlinks they offer to Web users.

The current three colour set may not be suitable for certain colour-blind audiences, such as tritonopes. The design can be experimented with using any other three colours, preferably from the basic colour set, for the broadest group of users.

4.5.3 Application of the Product - Examples

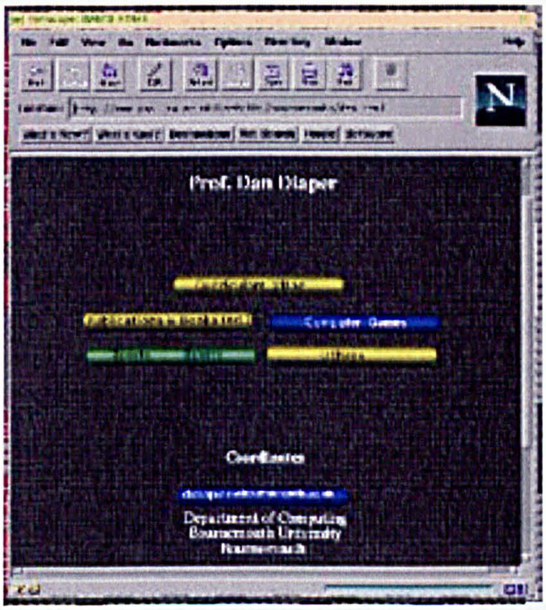


Figure 4.5: A screen dump of a Web page (previously shown in Figure 4.1) with an application of the Web product i.e. the three colour scheme. The plate gives a user, at a glance, consistent information concerning screen areas that are selectable (yellow) and not selectable (blue).

Figure 4.5 illustrates a Web display generated after applying the three colour scheme to the display in Figure 4.1 which did not follow the design. Figure 4.1 uses a single colour consistently across the display, however, it provides little information about selectable areas and associated hyperlinked documents when compared to the design illustrated in Figure 4.5.

Figure 4.5 illustrates some immediate advantages of the three colour scheme. For example, the buttons immediately allow users who know the colour conventions to distinguish between quickly accessible areas (yellow) from the areas with medium accessibility (green) or low accessibility (blue). The desirability is assumed to be uniform for all the links.

It was found relatively easy to use Java button applets and image maps in place of existing hyperlink locations. This allowed the research to design some experimental pages for the purpose of testing the validity of the colour scheme by evaluating end users' performance illustrated later in Chapters 9, 10, 11 and 12.

4.6 In Retrospect: Guidelines for Web Designers

One can divide Web page design into three issues. First, the information that is to be conveyed to people needs to be generated and structured. Second, the hypertext based methods, often more than one (i.e. many hyperlink destinations have many-to-one hyperlinks), by which users access the information need to be developed. Third, organising the hyperlinks in order to reduce the search time.

For all but the simplest of Web sites, there are usually a number of different ways to organise information and provide mechanisms to navigate through it. Once the general categories are established, then any available display features can be used to signal levels of importance. Designers may consider the following guidelines while organising the Web sources:

- Visual signals that provide feedback from the display are fundamental design features which are often overlooked or under exploited. Use available, inexpensive display features (for example, colour) that reveal the site or document organisation and nature of the linked-to resources.
- Since users can become easily bored or irritated with slow and undesirable links, and have difficulty finding specific information quickly among available links, develop useful classifications to support scanning and locating different categories of resources available, visually *at a glance*.

Many sites attempt to attract Web users with large or many graphics, audio and video, forgetting that it is not sufficient for a Web page to offer merely a visual experience. The functionality with respect to accessing information is very important in order to retain its visual appeal [Brown 1996]. A complex home page that takes forever to download and does not fit on the average reader's screen, and offers little or no functionality, will be unsatisfactory to most Web users. For example, pages with unwanted advertisements and data dumps, and also pages with a sequence of forms that force the user to enter lots of information. Design solutions that track down desirable data reduce user effort and time, and consequently increase user satisfaction.

Chapter 5

Introduction to Usability Studies

5.1 Empirical Approach to Evaluate the Colour Designs

The colour designs proposed in sections 3.5.1 and 4.5 represent the pilot and main products of this research on the traditional UI and the Web platforms, respectively. While the delivery requirements play a crucial role in the colour designs, they also drive the evaluation of the validity of the designs by adopting a careful empirical approach. The usability studies reported in this thesis were intended to meet the requirements generated from considerations about how to deliver the results of the research to the software industry as well as the requirements of the product design. Further, the empirical approach provides a rich, evaluative function of both the colour design(s), and the proposed means of delivery.

Several experts have described that the three essential features of the field of HCI are (i) know the user, (ii) know the task, and (iii) know the user environment [Faulkner 2000]. In the process of designing and delivering computerised systems, usability engineering attempts to address the three issues mainly from the practical end. One important aspect of the empirical approach adopted in current usability studies was to examine users' ability to learn (based on specified amount of training), remember and apply (see sections 2.3 and 2.4) the product design both within and across sessions, and applications. An attempt has been made to obtain empirical evidence on how the designs help and hinder users in work environments [Landauer 1993]. The studies used an experimental method derived from psychology [Greene and d'Oliveira 1978, Johnson and Solso 1978] and were initially aimed at investigating a set of conditions that demonstrate the sensitivity of the proposed test measures.

Overall, the usability studies attempt to investigate whether it is possible to use the visual feature, colour, in the manner it is proposed in this thesis. The experimental studies attempt to apply the colour designs to the existing application interfaces and evaluate their effect on users' performance. It should be noted that exploring other information coding techniques is beyond the scope of this research.

5.1.1 Setting Up the Testing Environment

Faulkner [Faulkner 2000] says -

"It is important for the system's designer to have a clear idea of the tasks the users do and the environment in which they carry out these tasks. In ergonomics the environment in which the task takes place is one of the paramount considerations ..."

Current literature on experimental studies describes different types of testing environments [Faulkner 2000, Nielsen 2000]. Some usability studies are conducted in controlled laboratory environments where parameters such as lighting conditions, temperature, working distance are always maintained constant and the user testing is monitored by using sophisticated equipment such as multiple video cameras to study different aspects/angles of the test environment. Usability experts have always questioned the use of such controlled, expensive and laborious testing environments in evaluating the usability of a product with respect to real user working environments where the previously mentioned parameters vary to a large extent. Consequently, there exist some recommendations (i.e. discount usability studies) to design and conduct studies in real user working environments without extensive use of equipment [Nielsen 1993, Nielsen 2000].

This thesis takes a view that if delivered, the proposed colour designs will be used by the people in normal working environments where the lighting conditions may be poor, monitors are generally adjusted to users' comfort levels etc. Consequently, the user testing has been planned to take place in a real working environment (i.e. a small computer office). Some adjustments were made for maximising user comfort. For example, the monitor and the user were positioned in a comfortable location in a computer room in order to minimise the negative effects of lighting such as glare. Distance between the user and the monitor was adjusted to meet users' comfort levels, and contrast and brightness of the monitor were adjusted suitably to provide optimum viewing conditions (with respect to ambient light) to clearly see the colours by the users.

Further the studies minimised the use of equipment in that sophisticated equipment such as ambient light sensors was not used since the studies were not planned to create an idealised work environment for the user. The studies mainly intend to examine how the users perform a specified set of test tasks. Analysis of the other parameters such as user's emotions etc. are not central to the studies. Consequently, from the task analysis standpoint, it was decided to only monitor user activities in performing the specified task(s) using one video camera. This arrangement reduced the amount of work involved in creating the testing environment.

In summary, there is a great amount of diversity in real work environments. Parameters such as display typeface, luminance levels, viewing distance, ambient conditions etc. are extremely sensitive to small values. As a result, the size of effects of these parameters were not considered for analysis in this research. The studies were directed at displays normally employed in real offices.

5.1.1.1 Recruiting Users:

Throughout the research, recruiting test users was found very time consuming. Notices requesting user participation were displayed on the notice boards. Further, test users were contacted through e-mails, friends and colleagues. Based on mutual convenience, time slots were arranged for each user. During the allocated time slot, each user was asked about their colour vision, computer skills, and willingness to complete the experiment. For the first three experiments, test users were required to have skills in using computer systems and to be able to navigate through applications by using mouse and keyboard. For the Web experiments, the users were required know how to use the Web. A minimum of three months experience in using computer applications and/or the Web was considered as the basic guideline in recruiting the users. They were aged between 18 - 40.

Further, it was decided to recruit mainly Liverpool University students for two reasons - a) availability and b) computer skills. Also, the research did not primarily intend to recruit professional graphic designers for the Web experiment, because an average Web author who designs Web pages is not an expert. It follows the delivery requirements (see Chapter 4) of the Web colour design in that it is intended to be used casually by an average Web author.

In studies that required multiple user groups, it was necessary to adopt a strategy to allocate users. This research adopted one of the powerful and widely used strategies namely, random allocation [Altman 1991, Jadad 2001]. User names were picked up at random from the list and allocated in equal numbers irrespective of age, sex and computer skills.

Prior to commencing the experiment, it was made sure that everything was in full working order. For example, a properly working computer and its accessories, applications in use, Web browser, Web pages etc. Following this, the user was invited in and introduced to the computer and test environment in general. It was made sure that the user was fully comfortable with seating, the positioning of keyboard, mouse, monitor, brightness etc. The user was made aware of the overall purpose of the experiment - exploring the uses of colour as an important visual cue. Along with some instructions, handouts containing the details of the training and test tasks were given to each user. The user was also informed that they could terminate the experiment at any time. Following this, the investigation began by asking the user about their colour vision. Further details of the individual experimental method are given in Chapters 6 to 12. When the user had completed the test, their time and effort was appreciated and rewarded.

5.1.1.2 Testing Colour Deficiency

As part of recruiting, users were tested for normal colour vision[1] by employing the concept of knowledge elicitation [Diaper 1989]. During the introductory stage, each user was shown three screen displays that implemented the proposed set of colours (see Figure 3.1 for an example display). Information about the user's colour vision was derived from their replies to questions such as "could you please identify the colours on the display", "are you able to distinguish all the colours on the display?" The screen displays used one of the 20 plates of Ishihara test (see section D.3) just for the purpose of knowledge elicitation.

The procedure employed in determining the normal colour vision was found adequate as it successfully eliminated the participation of three colour deficient users in the entire series. Many users (who are mostly Liverpool University students) already had a thorough understanding of their colour vision. Once the study began, users with colour deficiency immediately informed the experimenter about their colour vision and consequently their participation was immediately discontinued.

Since the above procedure was adequate to identify users able to perform the specific experiment, the Ishihara test was not carried out in detail. This saved considerable time and effort in conducting the experimental studies. Following the concept of discount usability studies [Faulkner 2000, Nielsen 2000], where feasible the thesis recommends the use of such simple and efficient procedures. The results of such a procedure are of

[1]This research interprets the normal colour vision as users' ability to see and distinguish proposed set of colours in the colour designs.

course not to be considered a valid medical test for colour deficiency.

5.1.2 Usability Metrics

Usability metrics represent the measurements that can be taken to judge the usability of a system. The existing literature provides several attributes that describe usability and its associated metrics [9241-11 1997, Shackel 1986, Hix and Hartson 1993, Faulkner 2000]. For example, Shackel defines usability in terms of learnability, effectiveness, attitude and flexibility, and ISO describes it in terms of efficiency, effectiveness and user satisfaction. Example metrics included the time required to perform the task, the time spent in using on-line help etc.

In this research, the metrics mainly considered in evaluating the usability of the colour designs include:

- The time taken to complete the test task
- Number of non-optimal responses
- Number of actions in completing the test task
- Questionnaires that indicate subjective ratings of the designs

Experimental Chapters 6 to 12 include the details of the above metrics. The basic structure of each of the studies include identifying the target group, recruiting users, establishing the training and test tasks, carrying out the evaluation and reporting the findings.

5.1.3 Planning a Series of Usability Studies

Evaluating the Pilot Colour Design:

The first three studies aim at evaluating the pilot colour design and the grouping effects of colour on user performance on traditional UI platforms. The first experiment was planned to maximise the effects of consistent use of colour over its monochrome counterpart. Once benefits of using colour were established, the second experiment was intended to find the detrimental effects of inconsistent use of colours over consistent usage (both colour and monochrome). Both the studies were mainly aimed at finding the beneficial impact of the pilot colour scheme on user performance. The third experiment was solely based on the outcome of the second experiment. Having not found significant performance improvements with the pilot colour scheme, the third study was conducted to verify the already established effects of crude colour grouping, for example reduction in overall search time which is highly desirable in search oriented tasks.

Besides the informal interviews, the metrics mentioned in the previous section played a major role in establishing the usability of the pilot colour scheme. The studies used different tasks and user groups under varying test conditions. The first two of the studies utilised software applications that were already available and the third used a task designed for the experimental purpose only. The details of the training and test tasks, recruitment of users and evaluation procedures are described in Chapters 6, 7 and 8. The hypotheses being examined in these experiments were:

1. It is possible to assign functionality to colour other than using it simply for crude grouping and aesthetic appeal in UI design.
2. People can associate a colour with a specific function.
3. It is possible for people to learn, remember and apply colour rules.
4. Proposed functionality of colour in traditional interface design is beneficial when applied consistently both within and across applications.
5. A consistently coloured interface has a beneficial effect on users' task performance over both monochrome and inconsistently coloured interfaces.

Overall, the first three experiments included two types of tasks:

- **Search-Identify-Select:** In this case, the test user searches for the target screen area, for example a pushbutton or a text entry box. Once identified, the user simply selects the appropriate target object or performs the required operation, such as text entry, as per the given instructions. Test users who know the colour scheme will apply the rules while performing the task.
- **Subjective Ratings:** Test users were asked questions about the colour rules and the study in general, and were asked to rate their opinions on a 7-point scale. Some questions were concerned with users' ability to retain colour rules over a period of time.

All the three experiments included on-screen displays of applications that implemented the colour design. Users' task performances were video taped in order to perform a detailed analysis of data at a later stage.

Testing the Web Colour Design:

The first three studies formed the basis for designing the studies that evaluated the main product of the research - the Web colour scheme. Having successfully tested the colour grouping effect in Experiment 3, the research focuses on establishing the same conclusions for the Web colour scheme in terms of achieving reduced search times for any Web user, *at a glance*. The failure of the generality of the pilot colour scheme in Experiment 2 has raised some serious concerns about the empirical approach to be taken in designing and validating the Web product. The empirical studies described in Chapters 9, 10, 11 and 12 provide a rich, evaluative function of the Web product and test a number of different issues. In particular, they attempt to examine the Web scheme in terms of better user performance and satisfaction obtained from using colours that provide some information about Web. The hypotheses being examined in these experiments were:

1. It is possible to assign novel functionality to colour in WPD other than for crude grouping and aesthetic appeal.
2. People can associate a colour with a specific function. The function could be complex.
3. It is possible for people to learn, remember and apply colour rules.
4. Colour learning can over-ride users' biases, such as left-right selection bias.
5. The proposed functionality of colour in WPD is beneficial when applied consistently both within and across applications.
6. Generality is universal: The proposed Web product can be applied to any page already available on the Web without affecting the original contents.
7. The Web product can be easily implemented by Web authors.

The test tasks were similar to those in the first three experiments:

- **Search-Identify-Select:** In this case, the test user searches for the target link area, for example a button or an image map. Once identified, the user simply selects the appropriate target link or performs the required operation, such as text entry, as per the given instructions. Test users who know the colour rule(s) will apply them while performing the task.

- **Question Answering:** This common technique was employed [Ringel and Hammer 1964, Tullis 1981] as part of the test task design. The questions used were of simple fact-retrieval type. For example, "Where did Dr. Diaper work in 1990?" The task required the participants to read the on-line text in order to find the correct answer. This was very straightforward in that the user was not required to interpret the text to answer the questions. The technique was used in some of the experiments in the current series (Chapters 9 to 12).

- **Subjective Ratings:** Users satisfaction was described as a very complex attribute in determining the usability of a product/system [Faulkner 2000]. In this research, test users were asked simple questions about the colour rules and the study in general, and were asked to rate them on a 7-point scale. Some questions were concerned about the users' ability to understand and retain the colour mappings over a period of time.

To be more realistic, all of the Web studies involved on-screen display of the colour design implementation instead of a hard copy of screen dump. Experiment 7 was an exception in that paper based screen dumps of displays were used in the early phases. In contrast to the first three studies, users' task performances were not video taped. Automatic logging procedures [Nielsen 1993, Faulkner 2000] were used in order to perform a detailed analysis of data at a later stage. A questionnaire (see Chapter 12) was used in order to obtain data regarding user evaluation of the Web product.

Equipment Used:

All the studies used the same machine: Hewlett Packard Unix machine (product code: HP A1097A; HP model 710 workstation), 19 inch monitor, 1280 x 1024 resolution, Graphics option: 19-inch 1280x1024, 8 plane colour model. The monitor was set to display colours in all testing conditions that used the colour design. In the first three experiments the RGB values of colours used were: light blue (173, 223, 255), light green (65, 163, 23), red (255, 0, 0), wheat (243, 218, 169), white (255, 255, 255). In monochrome testing conditions, different shades of grey were used instead of the above shades of blue, green, red and yellow. The lightness of colours were nearly preserved in attempt to provide optimal viewing conditions. The associated RGB values of grey shades were (215, 214, 213), (94, 95, 92), (127, 127, 127) and (206, 208, 204), respectively. The third experiment in the series (see Chapter 8) has also used a shade of pink with values (196,129,137). The Web experiments used three shades of blue, green, yellow with RGB values (0, 0, 153), (0, 102, 0), (255, 255, 102), respectively from the Web safe colour palette[2] in order to achieve solid, non-dithered, and consistent displays on any computer monitor, or Web browser, capable of displaying at least 8-bit colour (256 colours). Throughout this thesis, these different shades of colours are simply referred as blue, green, yellow, red, white and pink. The choice of selection of these colours was not evaluated in depth in terms of determining any other visual effects that users might suffer, for example after images, depth effects and fatigue.

Room Layout and Viewing Environment Parameters:

The viewing environment parameters (with approximate values) include:

(1) Eye-to-screen distance: 50 to 60 cms,

(2) Vertical Location: Viewing area of the monitor between 25 to 30 degrees below horizontal eye level,

(3) Monitor Tilt: Top of the monitor slightly farther from the eyes than the bottom of the monitor,

(4) Lighting: Outside light was controlled with blinds; Illuminance levels around 300 Lux,

(5) Colour Temperature: In the range of 5000 - 5500K,

(6) White Point: Set to 6500K D65 standard; Luminance level - at least 80 candelas per square meter,

(7) Ambient Light Sources: Mainly normal window light and fluorescent light. Any sources of reflection or glare were avoided, and

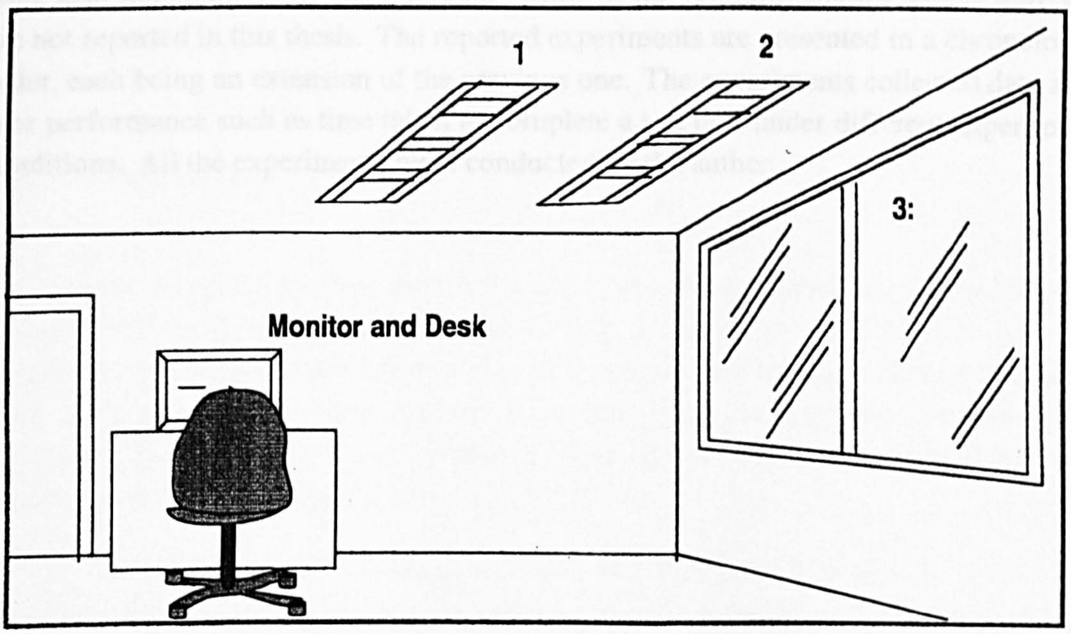
(8) Level of ambient illumination: Less than 64 Lux when measured in any plane between the monitor and observer.

---

[2]While using the Web safe palette, the thesis acknowledges the associated strengths and weaknesses discussed in depth by several graphic design experts [Priester 2001]. For example, if a Web author is designing for the mass market, including persons with antiquated computer systems and slow modems, then it is recommended to be Web Safe. On the other hand, if the audience are intended to be computer specialists who have more modern displays and faster modems, then the design requirements change with the availability of alternative palettes such as Adaptive and Web-Snap Adaptive, and other adaptations of the 256-color palette.

The following diagram shows the room layout along with the sources of ambient light used in the experiments.

Experimental Room Layout



Main Sources of Ambient Light:

1, 2: Fluorescent Lights

3: Window
- **parallel to light features and at right angle to screen face**
- **window is about 20 feet away from terminal**

Test users[3] with different academic backgrounds were invited to perform a few well-chosen tasks and to "think aloud" [Diaper 1989, Nielsen 1993] about what they are doing. The IT experience of users varied from a minimum of three months to four years. The test tasks were performed by one user at a time to minimise differences in system response times. In a majority of situations, the system load (i.e. number of users logged onto the system at any time) was maintained relatively constant. Since all the Web pages designed for the experiments were published on the local Web site, downloading times were also maintained low and constant. Some of the pre-pilot studies which were tried are not reported in this thesis. The reported experiments are presented in a chronological order, each being an extension of the previous one. The experiments collected data about user performance such as time taken to complete a test task under different experimental conditions. All the experiments were conducted by the author.

[3]Following Nielsen's terminology [Nielsen 1993], the word "test user" was used throughout this thesis to refer to each person who participated in any of the experiments.

Chapter 6

Experiment 1: Colour Vs. Monochrome

6.1 Foreword

This chapter discusses the first usability study in the current series of experiments. The study implemented some real applications using the proposed UI colour design. The implementation included modifying the colour values of all interface objects and screen areas, both within and across sessions, to comply with the proposed five colours and associated functions. That is, each interface object or screen area was assigned one of the five colours based on its input focus.

Overall, the study comprised two parts. The first part served as a pilot study. Data collected in this part was analysed in order to test the sensitivity of test measures (see section 6.2.1). Once the sensitivity of test measures was established, the second part was carried out with some more test users. Both the parts share identical design features such as the experimental method, materials and test tasks. Data collected from both the parts were combined and analysed to determine colour design effect versus the monochrome effect (see section 6.2.1). Not all the possible avenues of research leading from this initial study such as an evaluation of task complexity and type, and optimal colour choices are explored as the main purpose of these early experiments was to facilitate the design of the subsequent tests of colour functionality.

6.2 Design

The study attempts to compare the test measures for two different test conditions: 1) a colour condition and 2) a monochrome condition. In the former, the experimental tasks are implemented using the pilot colour design on a colour display. In the latter, the same tasks are carried out on a monochrome display.

Hypothesis:

The study was carried out with the hypothesis that: (a) the experimental design can detect differences in test measures, (b) there exists a difference between colour and monochrome conditions, (c) the shift from colour to monochrome will produce the biggest possible colour effect between the two and (d) users in the colour condition will perform better than those in the monochrome in terms of task completion times.

The study made use of two software tools, the 'Specific Action/Object Selector' (SASO) and the 'Task Descriptive Hierarchy Editor' (TDH) which are part of the LUTAKD (Liverpool University Task Analysis for Knowledge Descriptions) tool, developed at the University of Liverpool as a part of TOM[1] project. The first tool, SASO (See D.1) takes one or more prose descriptions of the task as input and provides a suitable environment for the user to construct and manipulate lists of the objects and/or actions involved through the different command buttons. The task to be performed using this tool involved constructing a list of specific objects, for example animals and birds, from a prose text on living animals [Simpson 1961, Porter 1967]. The second tool, TDH (See D.2) allows both the construction and manipulation of hierarchy like structures that describe a task in terms of objects, actions and the relationships between them. The associated task in this experiment involved constructing a hierarchy of living things from the previous list of animals and birds.

The implementation of the front end of the experiment comprised all the input styles and their corresponding colours as proposed in the pilot colour design for traditional user interfaces (see section 3.2).

6.2.1 Design Rationale

Some reasons for using the LUTAKD toolkit in designing test tasks include:

[1]The TOM project (Task oriented Modelling for Interactive System Design) is a collaboration between Logica, British Aerospace, University College London, Queen Mary and Westfield College, and the University of Liverpool. It is led by Logica Cambridge Ltd and is funded under the DTI's Information Engineering Advanced Technology Program (Grant no. IED/4/1717).

- Generality of use: One of the important requirements of the design and delivery of the pilot colour scheme is a wide application range. This study intends to use real applications in order to demonstrate generality of its usage with any software application or a toolkit such as LUTAKD.
- Availability of LUTAKD source code: colour values of the interface areas and objects could be easily changed by directly modifying the source code
- Task environment initiates transfer: the toolkit comprises multiple tools related to each other. It is therefore possible to build a work environment where users can learn and transfer their experience from one to another.

The purpose of this study is twofold: Firstly, the study aims at carrying out an initial testing of the sensitivity of the measures of test users' task performance. The reason is to examine whether the experimental design detects any differences between the measures. If the design is insensitive to the measures, then it may damage the whole purpose of conducting the experiment.

Secondly, once the sensitivity of metrics is established, the study intended to find the colour design effect versus monochrome effect, for example in terms of task performance times and non-optimal response counts. As mentioned in the literature review in Chapter 2, some of the previous empirical studies did not report any significant differences in users' task performances between colour and monochrome displays [Travis 1990, Munns 1968, Jeffrey and Beck 1972, Cakir *et al.* 1980]. The experiment was intended to obtain some results in this direction. If the study fails to determine any effect, then it is possible that the metrics used in this research are not good for their intended purposes.

Of special interest to the general research is the question: 'should the colour design be made implicit or explicit to the users in the instructions?' The answer to this question is important because of the delivery requirements. After some deliberation, it was finally decided to consider explicit design alone in this study since one of the requirements of delivery is to explicitly teach users the colour rules while they use the colour design implemented applications in the real world. The use of both the test conditions was postponed to later experiments (see Chapter 7) in the series.

The results of this study will form an appropriate basis for conducting further studies.

6.2.2 Method

The test users were randomly assigned to two groups, representing the two test conditions mentioned previously. One of the groups served as a 'control group' and the other as an 'experimental group'. Test users in the control group used identical colour design displays while performing both the tasks. Test users in the experimental group initially used colour design displays with SASO and then worked on monochrome displays with the TDH.

On any one particular condition in this experiment, a test user was required to perform two test tasks. The first task was performed by using identical colour displays by both the groups. It was therefore expected to produce same performance in each group. The second task was performed on one of two different displays and was expected to produce differences in task performance between groups. As mentioned in the beginning of section 6.2, the control group users who used uniform colour displays were expected to perform better relative to the experimental group users who were transferred to monochrome displays.

The colour design was explicitly described to ensure that the users gave conscious thought to the new functionality of colour. Test users were instructed to remember (i) the functionality of different background colours and (ii) changes in display conditions (i.e. colour to monochrome), if any.

6.2.3 Materials

Handouts consisting of details of the experiment were prepared for distribution among the users. The study used the same Unix terminal connected to the same server throughout all experimental phases. The Unix terminal was equipped with a keyboard and a mouse (see section 5.1.3.3).

Further, the study utilised (1) a video camera to record users' test task performances for the purpose of a detailed analysis of test measures such as task completion times, and (2) a 24 inch colour television and a video player to view the video tapes.

The target test tasks used in this study were as follows:

- The task to be performed using the tool SASO (see section D.1) involved constructing a list of animals and birds from a prose text on living things.

- The second task involved constructing a hierarchy of living things from the previous list of animals and birds using the TDH (see section D.2).

The following two subsections describe the visual details of application screens used in performing the test tasks:

6.2.3.1 Design of Test Task 1

Task 1 Interface:

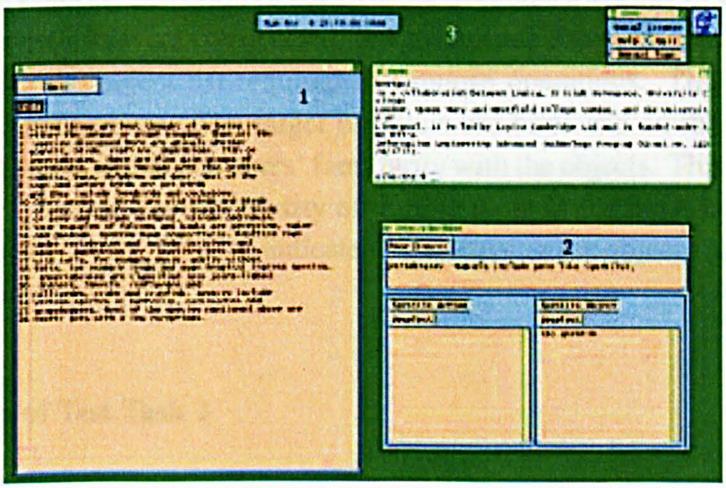


Figure 6.1: Experiment 1: A scaled down screen dump of the front end of Task 1 (SASO). The first application window (indicated by Number 1) contains the prose description of the test task (i.e. input text). The second (Number 2) window allows selection of objects from the input text in order to construct an object list. Number 3 indicates the root window as described in section 6.2.3.1.

The front end of the experiment that introduced Task 1 comprised three windows: one root window and two application windows (see Figure 6.1). The root window itself contained three hidden pop-up menus, one text window and six widgets, of which three were mouse selectable and the remaining non-selectable. The first application window displayed the prose description of the task, which included 18 objects to be selected through mouse operations. Further, the window contained four mouse selectable widgets. The second application window included six selectable widgets. The widgets used wheat i.e. light yellow (selectable), light blue (non-selectable) and white (keyboard input) background colours. The root window used green background colour (see section 5.1.3.3) as it contained hidden options. Users were instructed to construct an object list of 18 entries (see Appendix D.1).

Sequence of Operations in Performing Task 1:

Initially test users used a text window of the root window to execute SASO from the command prompt. The tool initially displayed the first application window which included eighteen target objects. Selection of any of these triggered a display of the second window from which object lists were constructed. Selecting each object from the prose description and adding it to the object list required five mouse operations. This number increases in the case of input errors. The target objects were birds and animals. The reason for choosing such target words was users' familiarity with the objects. This kind of familiarity was expected to reduce the complexity of finding them in the input text. The successful addition of objects to the list was indicated by enclosing the objects in round brackets as shown in Figure 6.2.

6.2.3.2 Design of Test Task 2

Task 2 Interface:

The front end that introduced Task 2 was similar to the one mentioned previously, except that it comprised three application windows (see Figure 6.3). The first application window included 24 mouse selectable widgets. Each of the second and third application windows contained one mouse selectable widget. The second window displayed the object list that had been constructed during Task 1. The third window was meant for listing selected actions, if any, from the input text in order to manipulate the object list. Action selection was not considered for the current experimental tasks and the third window remained unused without any action listing.

Sequence of Operations in Performing Task 2:

In Task 2, as in Task 1, test users executed the TDH at the command prompt from the text window. The TDH editor displayed three application windows. The first one allowed test users to construct a hierarchy of living things. The hierarchy included two elements: a class element and an object element. The first element referred to different classes

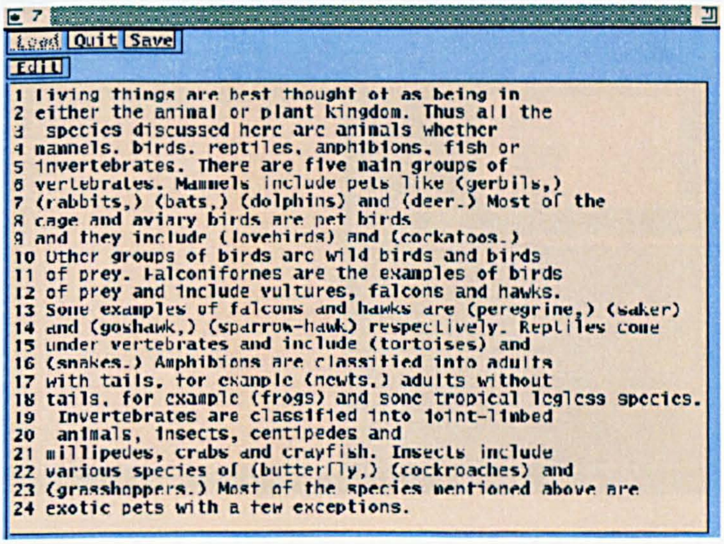


Figure 6.2: Experiment 1: A scaled down screen dump of an application window of SASO. As part of Test Task 1, users are instructed to construct an object list of 18 entries. The selected objects are enclosed in round brackets as shown in this figure.

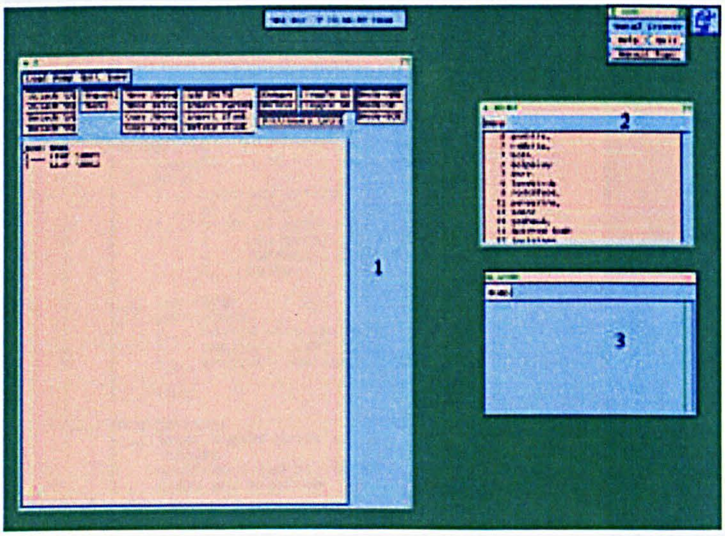


Figure 6.3: Experiment 1: A scaled down screen dump of the front end of Task 2. The first application window of TDH editor (indicated by Number 1) allows users to construct a hierarchy of objects. The second (i.e. Number 2) window displays the object list for the purpose of constructing the hierarchy. The third window (i.e. Number 3) was not used in the experiment.

of living things, for example mammals, and the second represented example objects of a specific class, for example rabbits and dolphins. The first window also displayed the starting entry for the hierarchy. Adding a class entry required a minimum of two mouse operations and a keyboard entry. Attaching an object to a class entry required three mouse selections. Overall, the task of constructing the hierarchy involved 23 class entries and 9 object entries out of 18 selected in the previous task (see Figure 6.4 and Appendix D).

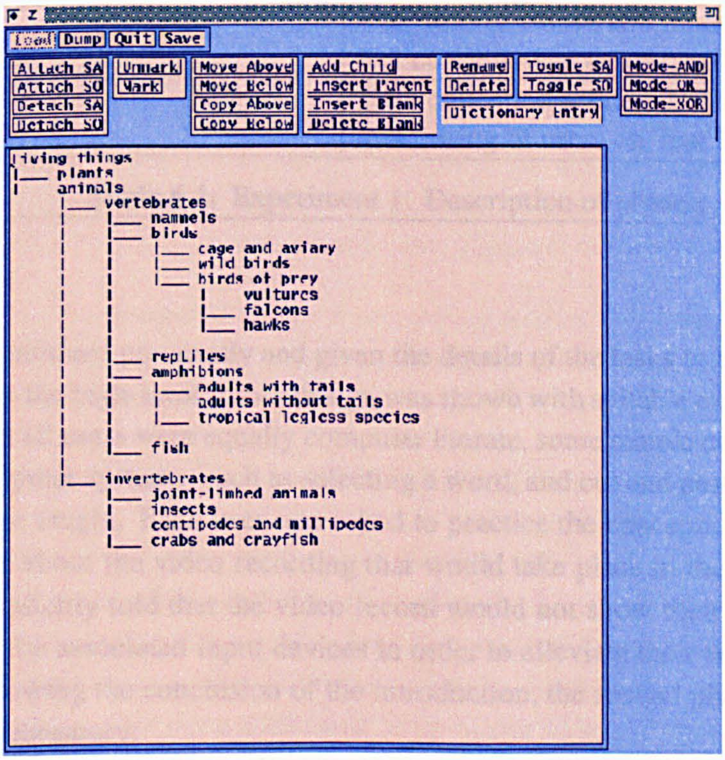


Figure 6.4: Experiment 1: The application window of TDH editor shows a partially constructed hierarchy of Test Task 2.

6.2.4 Testing Procedure

The experiment consisted of four phases as described in Table 6.1. The first phase started with enquiring the user for colour deficiency. Next, the handouts comprising a detailed explanation of the experiment were distributed (see Appendix D). The user was introduced

Experimental Phase	*Duration (Minutes)*	*Description*
1	10	Introduction to the study and test tasks
2	20	Data collection session 1: Training and testing of users on Test Task 1
3	5	Instruction period: Random allocation of test users to colour and monochrome test conditions in equal number.
4	20	Data collection session 2: Training and testing of users on Test Task 2

Table 6.1: Experiment 1: Description of phases.

to the tools mentioned previously and given the details of the tasks to be performed using the tools. Later the high-level colour design was shown with suitable examples (see Figure 3.1). Since not all users were equally computer literate, some simple concepts concerning the use of computer systems, such as selecting a word, and cut and paste operations using the mouse were taught. Each user was asked to practice the concepts. In addition, users were informed about the video recording that would take place in the following phases. Users were explicitly told that the video record would not show their faces but only the computer and the associated input devices in order to alleviate their worries about being recorded. Following the conclusion of the introduction, the second phase commenced in the computer laboratory.

The second phase consisted of training and testing. Initially the experimenter showed each user how to run the computer program used in conducting the pilot study. After giving a few instructions about the experimental task and prompting the user to input necessary information, such as typing the command in the text input area in order to execute the SASO, the front end program introduced Training Task 1 with the SASO selector. Users were trained on the use of the SASO in order to familiarise them with running of the tool before any performance measures were made. This session used displays that implemented the colour design. Following training, the testing session commenced and the user was instructed to execute Test Task 1 running on the same lines as Training Task 1. The user's performance was video recorded. Test Task 1 involved constructing a list of animals and birds from the given text description (see sections D.1.3 and D.1.4).

The third phase was actually a preparation for performing Test Task 2 in the fourth and the final phase. During this, the test users were randomly allocated to experimental and control groups in equal numbers (see section 5.1.1.1). The final phase was similar to the second phase except that the participants were trained and tested on the use of the TDH editor. The user was introduced to Training Task 2. While the training and testing of the experimental group was done on monochrome displays, the control group continued on the same colour displays as in Test Task 1. All users performed the test tasks using the same machine but with one of the two colour settings. That is, colour values were set to different shades of grey during the monochrome condition (see section 5.1.3.3) and were reset to colours as specified in the colour design during the colour condition. The users in the former group were briefed about the changes in the display conditions and Test Task 2 was carried out on the same lines as Test Task 1 as described earlier. Task 2 involved generating a hierarchy of birds and animals (see Figure D.1 and section D.2).

The above method was carried out with one user at a time. Following the completion of test tasks by all the users, the video data was analysed to determine test measures, such as overall task completion times.

The design therefore resulted in two user groups, one for each test condition: (1) execution of both the test tasks on colour displays with consistent colour design implementation, (2) execution of the first test task on the colour display and the second on the monochrome display. In case of (2), the testing sequence goes from colour to monochrome. That is, first the users were subjected to learn colour meanings and apply them while performing tasks. Following this, they were given monochrome displays where the colour learning did not play a role. This immediate shift from colour to monochrome was expected to produce some difference and the current study was specifically focused on obtaining the biggest possible difference from between the two (see section 6.2).

6.2.5 Experimental Design

The experiment was conducted in two parts under similar test conditions (i.e. colour versus monochrome). The first part was a pilot study. The video data of the first eight users was analysed in order to determine the total time taken by each user to complete the two test tasks. Average task completion times in each test condition were computed and subjected to statistical analysis in order to determine the sensitivity of test measures i.e. whether the design can detect differences between the completion times (see section 6.2.1). The results obtained from this pilot study, along with its conclusions, are summarised in Appendix C (see section C.1).

The second part was an extension of the first. Having successfully detected the differences in the task completion times, the video data of all the twenty users in both the parts was analysed together to determine multiple performance measures such as the number of operations to complete each test task (see section 6.2.7). The video was rerun many times to verify the accuracy of measures and also to note down several comments made by users during testing sessions, although subjective evaluation was not part of this study.

The results obtained from the video analysis and conclusions drawn from the statistical analysis are summarised in sections 6.2.7 and 6.2.8.

6.2.6 Test Users

Overall, twenty Liverpool University postgraduate students aged between 20-26, with normal colour vision, served as test users. Seven of them were not native speakers of English, none had knowledge of the use of colour in user interface design and none had participated in prior experiments testing colour design. Test users comprised thirteen males and seven females with ten users in each group. The IT experience of users varied from six months to three years.

6.2.7 Results

The three test measures used in the analysis were average task completion times, non-optimal response counts and number of operations to complete each test task. This section encapsulates basic findings of the statistical analysis related to these performance measures as described in the following sections:

Non-parametric statistical tests are generally used when the distribution of the data is indeterminate or does not follow normal distribution or the data cannot be reduced to normal because of small sample sizes. Further, these tests do not use the actual measurements of the observed data and generally do not use the variance as a central feature of the statistical test. The design of the current series of studies assumes the non-normal data for the above mentioned reasons (the results also show possible non-normal data) and consequently employs a non-parametric test 'Mann-Whitney U Test' for the purpose of statistical analysis. The test was used to determine whether there was any significant difference between the test measures of control and experimental conditions. The test is one-tailed because the experiment assumes that colour-to-colour condition would produce better performance than colour-to-monochrome condition (see experimental hypothesis, section 6.1).

6.2.7.1 Completion Times

The primary measure in this study was the time taken to complete the test tasks. Task completion time, measured in seconds, was defined as the interval between the start and end of each test task. The average task completion time was calculated for each task for a given test condition as shown in Figure 6.5. They are 237.67 (task1) and 631.91 (task2) seconds for the control group and 235.66 (task1) and 784.21 (task2) seconds for the experimental group, respectively.

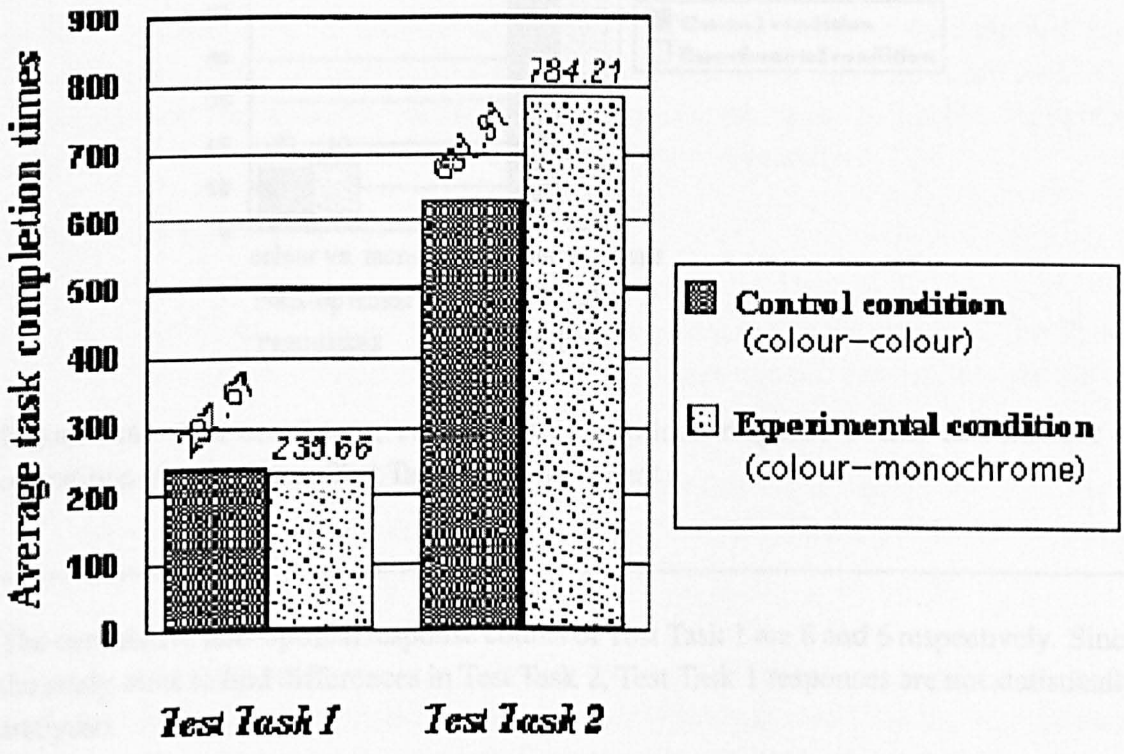


Figure 6.5: Test conditions and average task completion times in Experiment 1.

The Mann-Whitney U test results indicated a significant difference (Mann-Whitney U value of 4 and a one- tailed probability of 0.013) between the Test Task 2 time scores of the two test conditions.

6.2.7.2 Non-optimal Responses

A non-optimal response is said to have taken place when a user makes a wrong selection that does not produce the desired result. For example, if the user is expected to select a particular widget that results in a required action, but chooses a different widget that produces an undesirable or no action, it is counted as a non-optimal response.

The cumulative non-optimal response counts of Test Task 2 for the colour and monochrome conditions are 15 and 16 respectively as shown in Figure 6.6.

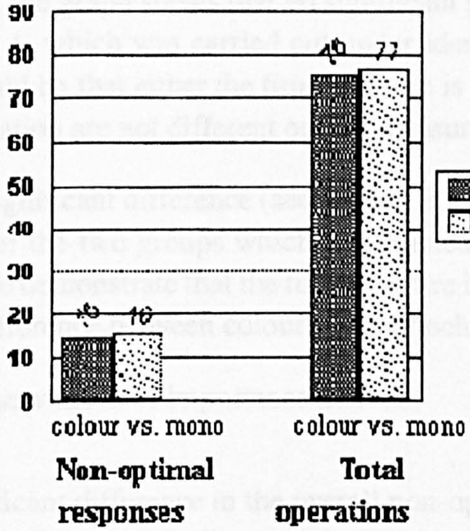


Figure 6.6: Test conditions, cumulative non-optimal response counts and number of operations in completing Test Task 2 in Experiment 1.

The cumulative non-optimal response counts of Test Task 1 are 8 and 6 respectively. Since the study aims to find differences in Test Task 2, Test Task 1 responses are not statistically analysed.

6.2.7.3 Number of Operations

Some examples of operations in performing test tasks include: selecting an object from the text displayed on the screen and adding an object to the object list. The average operation counts of Test Task 2 for the colour and monochrome conditions are 76 and 77 respectively as shown in Figure 6.6. Since the study aims to find differences in Test Task 2 only, Test Task 1 operation counts are not analysed.

6.2.8 Conclusions

The mean response times for each test task displayed in Figure 6.5 shows the correspondence between the colour and monochrome test performances. Even without applying statistical tests, the graph shows that no significant performance difference was observed with Test Task 1, which was carried out under identical conditions for both the groups. The reason could be that either the time measure is insensitive or the two samples drawn from the population are not different on this measure.

In contrast, a significant difference (see section 6.2.7.1) was found between the Test Task 2 time scores of the two groups which was carried out under different conditions. The results therefore demonstrate that the time measure is sensitive and is capable of detecting a significant difference between colour and monochrome conditions.

Some other observations of importance include:

- No significant difference in the overall non-optimal response counts of Test Task 2 in both conditions (refer Figure 6.6). That is, there was no speed-accuracy trade-off effect.
- No significant difference between the average number of operations occurred in completing Test Task 2 (refer Figure 6.6) in both the conditions, though they took different times (refer Figure 6.5).

Among the test measures, time is a continuous measure and non-optimal counts could be described as discrete [Diaper 1982]. The study demonstrates that the time measure is sensitive but not non-optimal counts. These results are consistent with the predictions made about these measures, that is, a colour effect is predicted primarily in terms of performance times (see section 6.2). The occurrence of a case where time is insensitive and non-optimal counts are sensitive would have been a major concern to this study.

In summary, Test Task 2 completion times demonstrate that the participants under identical colour conditions (control group) took less time and therefore performed better than those under colour-monochrome conditions (experimental group). This could be due to the overall positive effect of consistent colour functionality both within and across the tools used in this study. In other words, the change in working conditions from colour design displays to monochrome displays adversely affected the performance of the experimental group as indicated by the significantly higher task completion times. The results are therefore consistent with the experimental hypothesis (section 6.2). Having obtained the colour versus monochrome effect, this study encourages further testing of the use of colour under other test conditions, for example consistent colour versus inconsistent colour (see section 7.1).

Chapter 7

Experiment 2: Consistency Vs. Inconsistency

7.1 Introduction

The primary purpose of Experiment 1 was to test the sensitivity of the metrics, which would be used in the subsequent investigations. The choice of a colour versus monochrome effect was to ensure that the measurable differences between the two conditions was as large as it was likely ever to be. Experiment 2 was intended to approximately simulate how a real software designer might apply the colour design to the user interface in a simple, straightforward way.

One important aspect of the current study is to determine the scope of generality of the UI colour design. This is an important issue since the generality is driven by delivery requirements in that the designers who prefer to use the colour design would take a simple, easy approach while implementing it in any application. Other delivery requirements described previously (see sections 1.4.2 and 3.2), for example simplicity, consistency and explicitly teaching the rules, remain appropriate. Furthermore, the generality of the product drives the experimental design such that the method employed will evaluate whether the product can be applied to a wide range of applications.

Experiment 2 was designed to examine the colour design in more detail than the previous study. This experiment differs from the previous one in several aspects such as test users, tools and test tasks, front end design, the design phases and the number of test conditions. An important difference was the inclusion of an 'implicit test condition' in which colour functionality details were not described *a priori* to some test users. It was intended to test whether the colour design aided implicit learning of colour functionality. The following sections describe the method employed and present the results followed by conclusions.

7.2 Design

The study attempts to examine the test measures between the four different test conditions as shown in Table 7.1. The first two conditions used colour displays with the colour design implemented. The colour design was made explicit to the users in the first condition and implicit in the second. The third condition used two different displays. One display implemented colour design and the other inconsistent colours (it should be recalled that a colour design means using a consistent colour scheme both within and across applications). In the latter, colours were made deliberately inconsistent in order to obtain an inconsistent effect. The last condition implemented only monochrome displays. The interfaces had the same functional capability and low-level interactions and style in all four conditions. The purpose of the experiment was to fulfil the criteria listed earlier in section 5.1.3.

Hypothesis:

The study was carried out with the hypothesis that: (i) the explicit colour group will perform better than the other groups, (ii) there will be a consistent colour vs. monochrome effect as in Experiment 1, (iii) the inconsistent colour group will produce adverse performance relative to other groups and (iv) monochrome displays will produce better performance than inconsistent displays (the reason being that monochrome itself produces consistent grey-scale patterns which is relatively superior to a multicolour, inconsistent scheme).

The study comprised two main tasks. Three X windows applications, Xaero, Xalarm and Xwais (See Appendix E) were used in building the tasks. The first task uses Xaero and the second uses the other two applications. Xaero allows test users to create animations of simple mechanical components. Xalarm triggers an alarm message at specified intervals of time. Xwais searches for documents on a particular topic specified by the user and displays the results as a list of available documents.

7.2.1 Design Rationale

The study has used different applications in designing the tasks in order to test the generality of the colour design. The tasks were designed to be at least at the same level of complexity as those in Experiment 1. The main reason for using these three X applications was the ability to change the colour values of the interface objects and screen areas.

User Group and Test Condition	*Design*	*Task1*	*Task2*
G0 (Control)	Explicit	Colour Design	Colour Design
G1 (Experimental)	Implicit	Colour Design	Colour Design
G2 (Experimental)	Implicit	Inconsistent Colours	Colour Design
G3 (Experimental)	Implicit	Monochrome	Monochrome

Table 7.1: Test conditions in Experiment 2

An important goal of this study was to examine the robustness of the colour design with real applications. A plausible approach is to evaluate the effect of the colour design with respect to multiple test conditions. The four test conditions of Table 7.1 were expected to provide an adequate basis for this. The strategy adopted in analysing test conditions is described later in section 7.2.5.

The monochrome condition was included in the design to replicate the colour consistency versus monochrome effect of Experiment 1. Furthermore, the study intends to compare this with the inconsistent condition to test whether monochrome displays result in better performance than the inconsistent colour displays (see sections 1.1 and 2.3).

The results of this study will evaluate the generality and consequently the robustness of the colour design in a range of applications.

7.2.2 Method

The test users were randomly assigned to four groups - G0, G1, G2 and G3 - representing the four test conditions. G0 served as the 'control group' and the others as 'experimental groups'.

All test users performed the same test tasks under one of the four test conditions. The first task, using the Xaero application, was performed under different colour display conditions (see Table 7.1). It was expected to result in different user performances. The second task, using the Xalarm and Xwais applications, was performed under consistent display conditions: colour design or monochrome. Task 2 performances were expected to be consistent with the predictions made regarding colour-to-colour and colour-to-monochrome effects carried over from Experiment 1.

Test users in the control condition were instructed to remember and apply the colour rules throughout the testing sessions.

7.2.3 Materials

The materials were similar to those described in Experiment 1 (see sections 6.2.3 and 5.1.3.3). Handouts consisting of details of the experiment were prepared for distribution amongst the users. The study used the same Unix terminal connected to the same server throughout all experimental phases. The Unix terminal was equipped with a keyboard and a mouse. Further, a video camera was used to record users' test task performances and a 24 inch colour television and a video player to view the video tapes.

The test tasks were as follows:

- The first task involved animating a mechanical object which used three components: sphere, cuboid and a cylinder and a spring with pre-defined specifications (see Appendix E.1).
- The second task (see Appendix E.2 and E.3) involved two main events: setting the alarm time to eight minutes past the current time for the next three days and listing the available documents on the 'Internet'. The tasks included menu selection, text editing and search procedures.

A front end (see Figure 7.1) was designed using UIMX, an interface prototyping tool on Unix system, to allow the test users to access the X applications with simple mouse clicks. The implementation of all three applications and the front end comprised of all the input styles proposed in the colour design.

The following subsections describe the visual details of application interfaces used in performing the test tasks.

7.2.3.1 The Common Front End

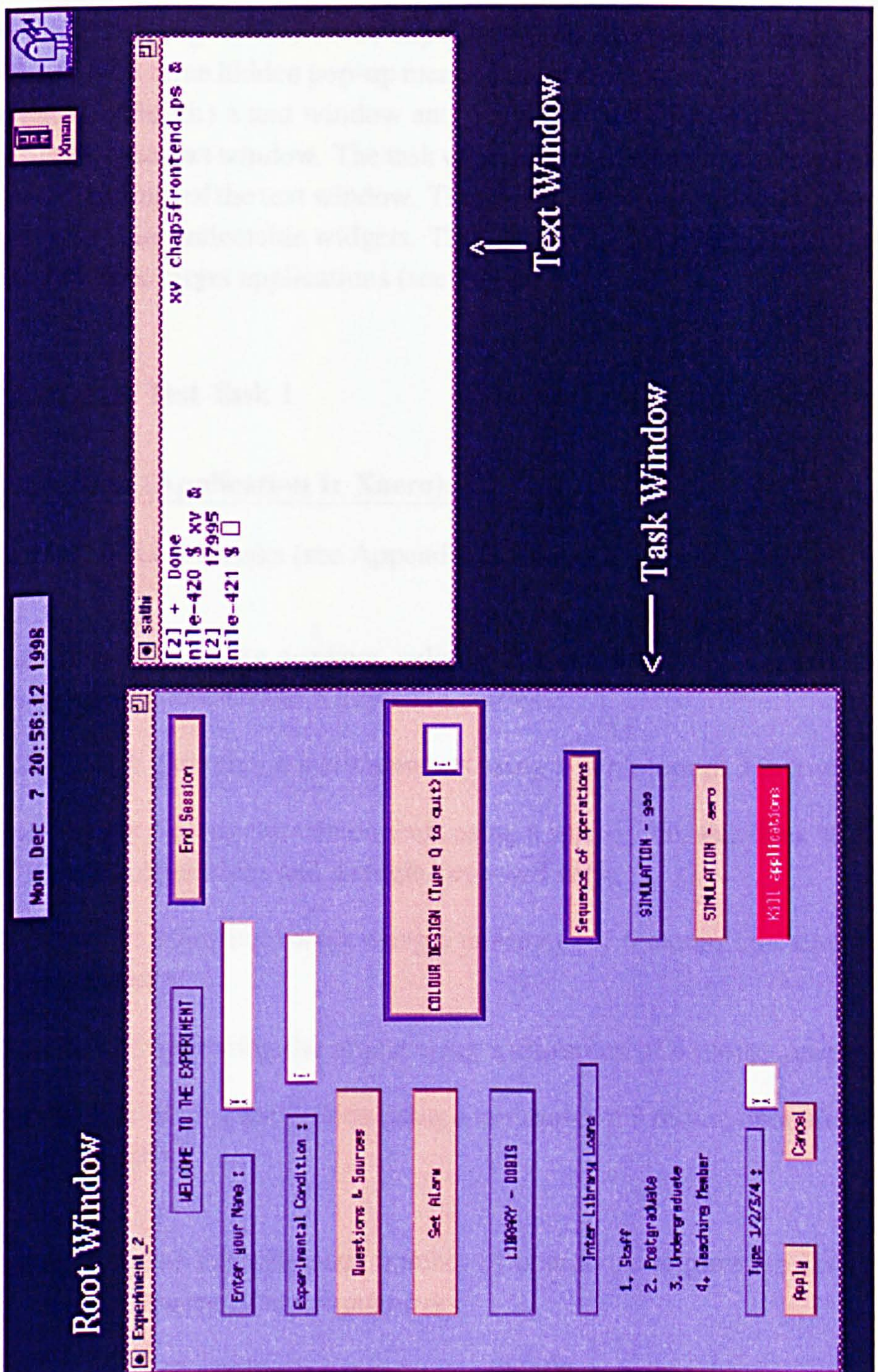


Figure 7.1: A scaled down screen dump of the front end of the test task design in Experiment 2.

The front end (see Figure 7.1) of the experiment was comprised of three windows: (i) A root window with three hidden pop-up menus and three widgets, one mouse selectable and two non-selectable, (ii) a text window and (iii) a task window. Participants used hidden menus to create the text window. The task window was displayed on executing a command at the system prompt of the text window. This task window contained 3 keyboard, 8 mouse selectable and 7 non-selectable widgets. Three of the mouse selectable widgets were used to invoke the three target applications (see Figure 7.2).

7.2.3.2 Design of Test Task 1

Task 1 Interface (Application 1: Xaero):

Task 1 included six sub-tasks (see Appendix E, Tables E.1 and E.2).

- Sub-Task 1: selecting a sphere, cylinder and cuboid which required a minimum of 18 mouse operations and 6 keyboard entries.
- Sub-Task 2: grouping the components using a minimum of 3 mouse operations.
- Sub-Task 3: linking the components using a spring; this was done with a minimum of 9 mouse operations and a single keyboard entry.
- Sub-Task 4: adjusting forces using a minimum of 6 mouse operations and a single keyboard entry.
- Sub-Task 5: animating the object using a minimum of 4 mouse operations.
- Sub-Task 6: exiting the session using a minimum of 9 mouse operations and a single keyboard entry.

The numbers indicate the minimum number of operations required to perform the tasks and may increase in case of any input errors.

Overall there were 12 application windows with 109 mouse selectable widgets, 49 non-selectable widgets, and 10 text widgets. The users were trained only on using 35 mouse selectable widgets and 7 text widgets that were required for performing the training task. Not all 12 application windows were active at the same time; only the windows required by any sub-task under consideration were visible at a given time.

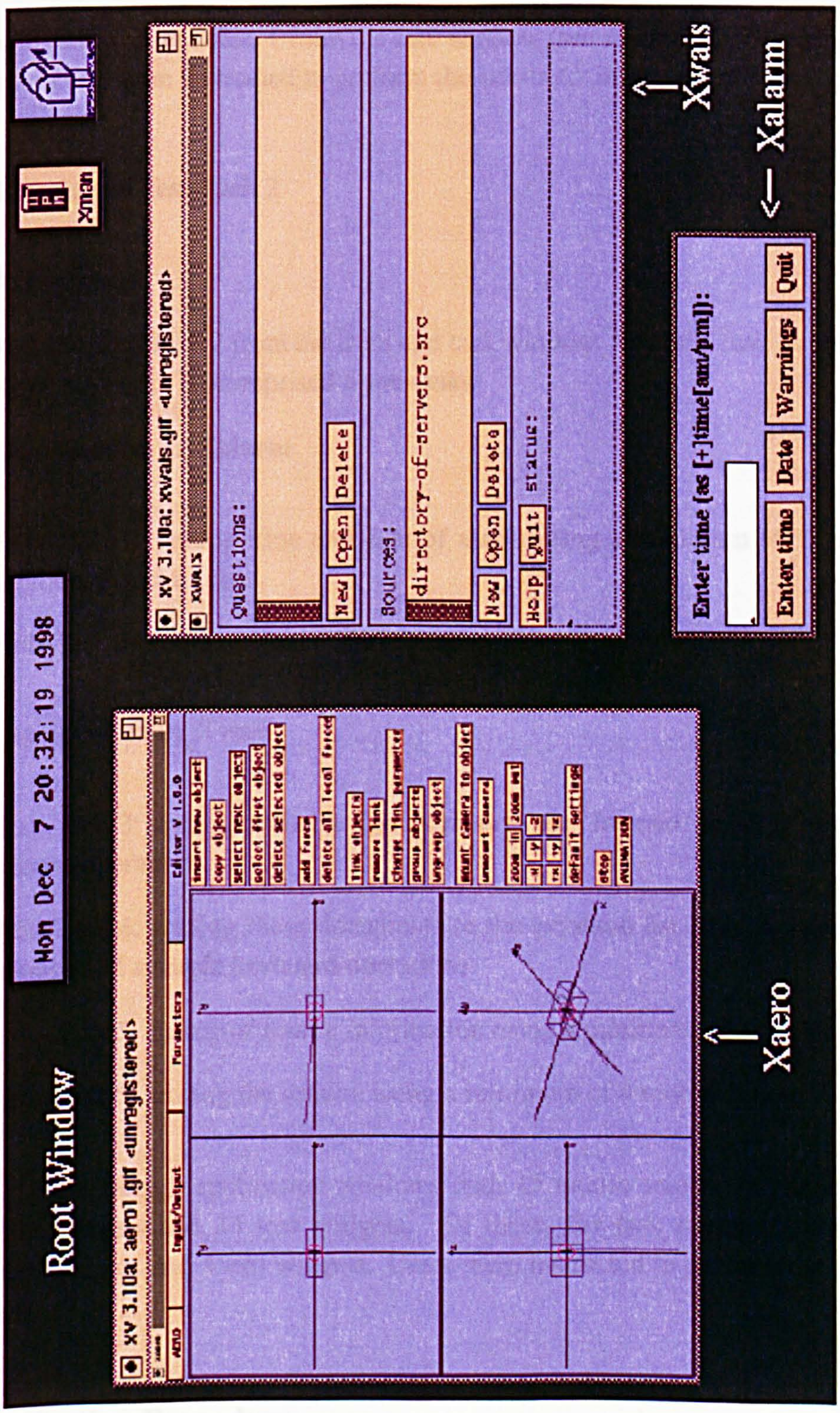


Figure 7.2: A scaled down screen dump showing three of the application windows of Xaero, Xwais and Xalarm in Experiment 2.

The test users executed Task 1 from the task window (see section 7.2.3.1) using the Xaero application and were instructed to perform the sub-tasks in sequence.

7.2.3.3 Design of Test Task 2

Task 2 Interface:

Test users invoked Task 2 from the front end task window. Overall, Task 2 (see Tables E.3 and E.4 in Appendix E) comprised 6 sub-tasks.

Using Application 2: Xalarm

- Sub-Task 1: setting time and date of alarm using a minimum of 3 mouse and 3 keyboard operations.
- Sub-Task 2: confirmation of alarm settings using a minimum of 4 mouse operations.

Using Application 3: Xwais

- Sub-Task 3: listing available documents on the 'Internet' using a minimum of 3 mouse operations.
- Sub-Task 4: adding these documents to the personal list using a minimum of 15 mouse and a single keyboard operations.
- Sub-Task 5: listing of source information using a minimum of 3 mouse operations.
- Sub-Task 6: exiting the session using a minimum of 4 mouse actions

Overall, there were 9 application windows with 45 mouse selectable widgets, 32 non-selectable widgets and 14 text widgets. Of these, the task required only 20 mouse selectable widgets and 4 text widgets. Users were instructed to perform the sub-tasks in sequence.

7.2.4 Testing Procedure

Experimental Phase	*Duration (Minutes)*	*Description*
1	10	Introduction to the study and test tasks
2	20	Data collection session 1: Training and testing of users on Test Task 1
3	20	Data collection session 2: Training and testing of users on Test Task 2

Table 7.2: Experiment 2 - Description of phases.

The experimental design replicated the overall design aspects and testing procedure of Experiment 1 except that there were only three phases in each of the four conditions as shown in Table 7.2. During the introductory period the users were introduced to the tasks described previously and given the details of the test tasks. As mentioned earlier, in the first condition the colour design was explicit. The test users in this condition were therefore instructed on the high-level colour design in detail. All the users were informed about some simple computer concepts such as selection and cut-and-paste of text using the mouse; they were required to practice them. The participants received handouts consisting details of the experiment.

During the first part of the second phase, test users were given training on Task 1. In the latter part, they were instructed to execute Test Task 1. The third phase was similar to the previous one. The participants were trained and tested on Test Task 2.

The users were tested on the same machine in both colour and monochrome conditions. For the monochrome condition, colours of the interface widgets were set to different shades of grey (see section 5.1.3.3). Different colours were used for the interfaces in the consistent (as specified in the colour design, see section 3.2) and inconsistent colour conditions (grey, pink, blue, white and red). Following the completion of the test tasks, users who were tested under colour conditions were informally asked to express their opinion on various aspects of the colour design (see section 7.3), although subjective evaluation was not included in the testing procedure.

The test was carried out with one user at a time and the user's performance was video recorded. The video data was later analysed to determine the test measures, such as overall task completion times.

In summary, there were four user groups and all of them performed two tasks under different experimental conditions as listed in Table 7.1.

7.2.5 Experimental Design

This section describes the strategy employed in analysing the test tasks under different test conditions. Following the hypothesis, the study attempts to analyse the test conditions / user groups in pairs i.e. G0 vs. G1, G0 vs. G2, G0 vs. G3 and G2 vs. G3 (see Tables 7.1 and 7.3) for the following reasons:

- **G0:** The study predicted that learning and consistency aspects within a task /application would influence users' performances across tasks / applications. In the first condition (G0), the colour design was applied consistently across both the tasks and was explicitly described to the users. The other conditions were designed to exclude either explicit learning or colour consistency or both across the tasks. The groups G1, G2 and G3, therefore, will be compared against G0 in order to measure the effect of colour. That is, G0 will serve as a control group to the other three conditions.

- **G0 vs. G1:** In the second condition (G1), the colour design was not explicitly described but applied consistently across the tasks. The effect of implicit learning of colour design could be tested in this condition. The implicit learning was predicted to take place during Task 1, which would influence the performance of Task 2. If there is no difference of performance between G1 and G0 in Task 2, then the implicit learning of colour rules could be considered as effective as the explicit learning.

- **G0 vs. G2:** In the third condition, the colours were completely inconsistent in Task 1, but consistent in Task 2. This shift was kept implicit, that is, not explained to the users. In other words, there was neither learning nor consistency of colour rules during Task 1. Comparison of the Task 1 performances with that of the control condition will provide a measure of any inconsistency effect.

- **G0 vs. G3:** As stated in the hypothesis (see section 7.2), the study attempts to determine colour versus monochrome effect. This can be achieved by analysing two possible pairs, i.e. G0 vs. G3 and G1 vs. G3. Since the study clearly intends to replicate the same effect of Experiment 1, only G0 vs. G3 pair is considered for analysis.

- **G3 vs. G2:** The study will compare these two groups (i.e. the third and the fourth test conditions) to test whether the monochrome condition results in better performance than the inconsistent colour condition.

7.2.6 Test Users

Twenty five Liverpool University students, aged between 18 to 27 years participated in this experiment. One user found the tasks very confusing and was unable to complete either of them. This user was therefore rejected. The other users did not report difficulties of this nature. All the users had normal colour vision and none had participated in prior experiments testing the colour design. Test users were randomly assigned to the four experimental conditions, with six in each. The IT experience of users varied from a few months to several years.

7.2.7 Results

The three test measures used in the analysis were average task completion times, non-optimal response counts and number of operations to complete each test task. This section summarises the statistical analysis related to these performance measures.

A Mann-Whitney U test is (see section 6.2.7) used to determine whether there was any significant difference between the test measures of the control and each of the experimental conditions. The test is one-tailed because the experiment predicted significantly higher performance in consistent colour-to-colour condition (the control condition) compared to consistent-to-inconsistent, colour-to-monochrome and inconsistent-to-monochrome conditions.

7.2.7.1 Completion Times

The primary measure in this study was the time taken to complete the test tasks in seconds (see section 6.3.7.1). The average task completion time was calculated for each task and a given test condition as shown in Figures 7.3 and 7.4. They are 422.2 (Explicit), 409.5 (Implicit), 471.5 (Inconsistent) and 404.1 (monochrome) seconds for Task 1, and 196.0 (Explicit), 181.3 (Implicit), 202.1 (Implicit) and 195.9 (monochrome) seconds for Task 2.

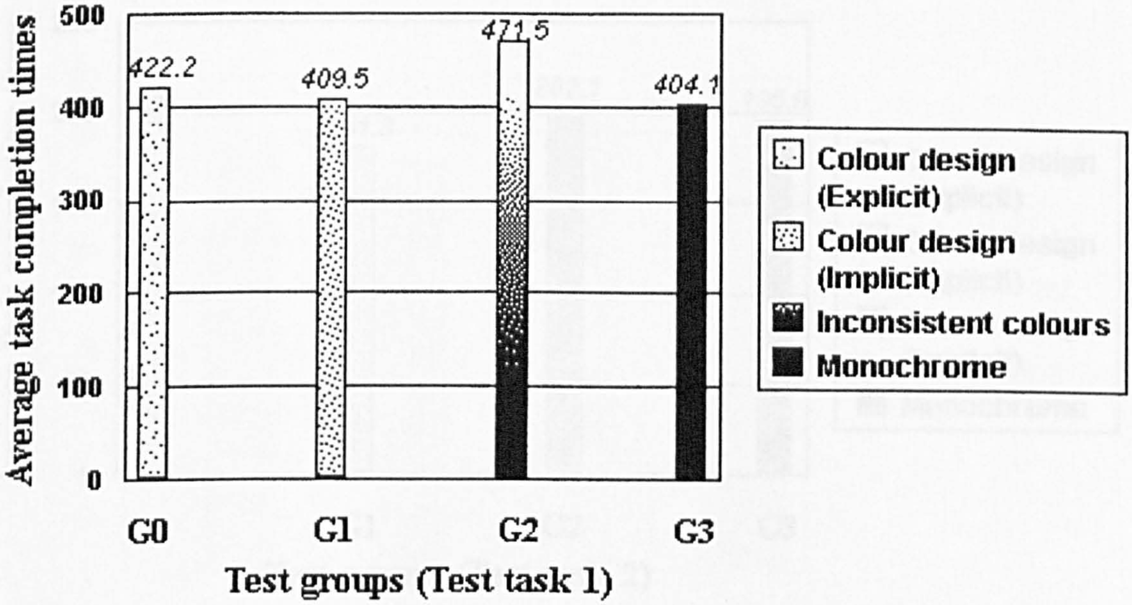


Figure 7.3: Test conditions and average task completion times for Test Task 1 in Experiment 2.

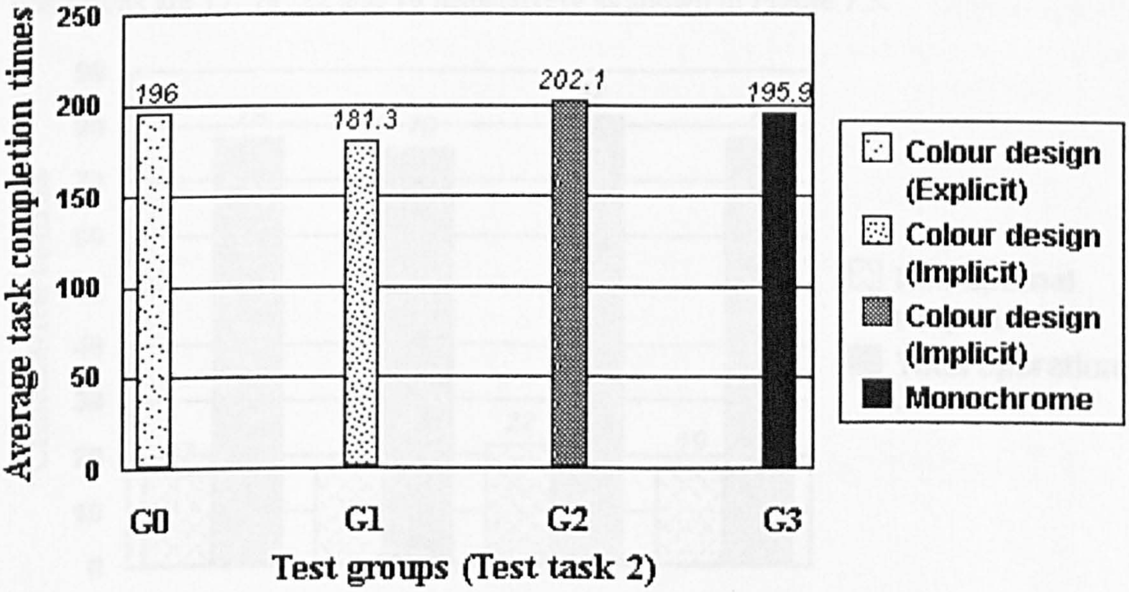


Figure 7.4: Test conditions and average completion times for Test Task 2 in Experiment 2.

The Mann-Whitney U test (see section 6.2.7) was applied to Test Task 1 completion times which was performed under different test conditions. The results of the test are shown in Table 7.3. The U values indicate no significant differences in the task completion times of all the four conditions.

7.2.7.2 Non-optimal Responses

The cumulative non-optimal response counts (see section 6.2.7) of Test Task 1 for the four conditions are 17, 18, 22 and 19 respectively as shown in Figure 7.5.

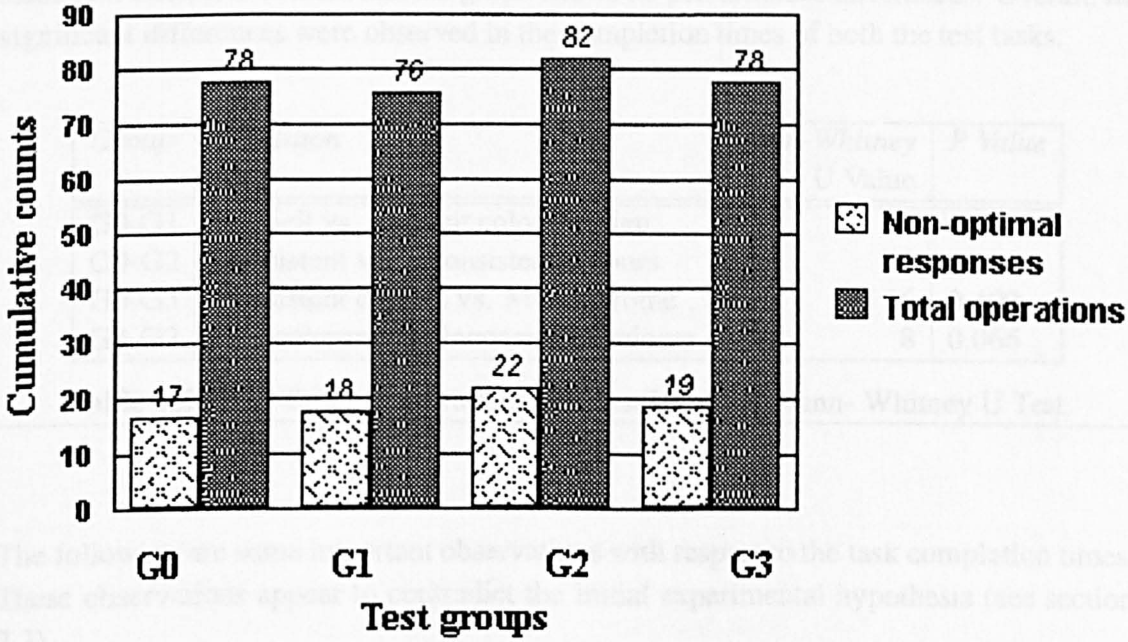


Figure 7.5: Test conditions, cumulative non-optimal response counts and the average number of operations in completing Test Task 1 in Experiment 2.

7.2.7.3 Number of Operations

Some examples of operations in performing test tasks include: selecting a spring from the options displayed on the screen and linking it to an object such as a sphere. The average number of operations for Test Task 1 for the four test conditions are 78, 76, 82 and 78 respectively as shown in Figure 7.5.

7.2.8 Conclusions

The mean response times for each test task displayed in Figures 7.3 and 7.4 shows the correspondence between the explicit, implicit, inconsistent and monochrome test performances. No significant differences were found in task completion times of all the four conditions (see section 7.2.7.1). In case of Test Task 2, even without applying statistical tests, it is evident that the graph shows no performance differences. Overall, no significant differences were observed in the completion times of both the test tasks.

Group	*Condition*	*Mann-Whitney* U Value	*P Value*
G0-G1	Explicit vs. Implicit colour design	14	0.294
G0-G2	Consistent vs. Inconsistent colours	10	0.120
G0-G3	Consistent colours vs. Monochrome	16	0.409
G3-G2	Monochrome vs. Inconsistent colours	8	0.066

Table 7.3: Test-Task1 (Experiment 2) : Results of the Mann- Whitney U Test

The following are some important observations with respect to the task completion times. These observations appear to contradict the initial experimental hypothesis (see section 7.2).

- No colour learning effect: There is no significant difference between performances of groups under explicit and implicit conditions. In other words, informing or not informing the users about the high level colour design appeared to have little effect as far as performance was considered. (G0-G1 in Table 7.3)
- There is no significant performance difference between groups of the consistent and inconsistent colour schemes (G0-G2 in Table 7.3). That is, the study failed to show significant performance gain due to consistent colour usage.
- No significant difference in average task completion times in colour vs. monochrome

conditions (G0-G3 in Table 7.3). This study therefore did not succeed in replicating the colour-monochrome effect observed in Experiment 1.

- No significant difference between the performances of groups under inconsistent and monochrome conditions (G3-G2 in Table 7.3). In other words, the study failed to show that monochrome displays are superior to inconsistently coloured displays.

Number	*Question*	*Response*
1	How do you feel working with consistent background colours across applications? *completely confused* — **1 2 3 4 5 6 7** — *everything made sense*	6.8
2	How sensible did the functional allocation of colour seem to you? *unacceptable* — **1 2 3 4 5 6 7** — *Excellent*	6.8
3	How easy was the colour design to learn and remember? *very difficult* — **1 2 3 4 5 6 7** — *very easy*	5.8
4	How easy it was to adopt to displays that implemented colour design? *very difficult* — **1 2 3 4 5 6 7** — *very easy*	5.8
5	How appropriately were the colours used? *very unsatisfied* — **1 2 3 4 5 6 7** — *very satisfied*	4.6
6	Do the inconsistently coloured displays affect performance? *Not at all* — **1 2 3 4 5 6 7** — *very much*	5.8
7	How comfortable you feel working with monochrome displays compared to colour? *feel completely frustrated* — **1 2 3 4 5 6 7** — *always know what link to select next*	2.2
8	Are monochrome displays better than inconsistently coloured displays? *Not at all* — **1 2 3 4 5 6 7** — *very much*	3.2

Table 7.4: Understanding of colour design - Summary of users' responses on a 7-point scale

---

The test users have produced some errors (i.e. non-optimal responses) by pressing the wrong command buttons. However, careful inspection of the data indicated no significant differences across the four conditions. Similarly, no significant differences were observed between the number of operations. As a result, no detailed statistical analysis was performed with respect to these two measures as the experiment was principally interested in examining the colour design effect on the task completion times.

A general observation of users' performance while executing the tasks, however, indicated that test users of consistent colour condition were less hesitant in performing the task than the users of the inconsistent condition. This could be due to the deleterious effect of the inconsistent use of colours.

An attempt has been made to identify any causes of the failure of the study. Appendix F.1 presents one possible analysis. No clear-cut reason however could be found. Time is potentially a sensitive metric and the task completion times indicate that the proposed colour design is clearly not universally applicable to any general application. In other words, as mentioned in the introduction, one of the delivery requirements that allow Web authors to casually apply the colour design has not be met. The product i.e. the colour design suffers from a potential delivery problem. This experiment was a failure in producing the expected effect of the colour design, which was based on the results of Experiment 1. To sum up, statistically, the study failed to prove the experimental hypothesis.

7.3 Users' Understanding of the UI Colour Design

The subjective evaluation was not part of the experimental design, however, following the completion of the experiment, the users were informally questioned about their understanding of the colour design. Since users were not tested on all conditions, first the author had made an attempt to explain all the test conditions involved in the study through some examples and the overall purpose of the study. They were asked to provide ratings for the questions listed in Table 7.4. Users' responses were scored on a 7 point scale with 1 being the lowest and 7 the highest.

Users shared nearly the same opinions regarding the colour design and consistent usage across applications. The ratings (see Table 7.4) indicate that the users prefer colour displays over monochrome. Users also favoured the consistently coloured displays as they found displays to be informative and user friendly. Comments made by the users during the experimental session also reflected such opinions. Overall, most of them appreciated the novel colour functionality and indicated the need for better functional designs involving the use of UI display features.

Chapter 8

Experiment 3: Crude Colour Grouping

8.1 Introduction

Experiment 2 has failed to establish the generality of the pilot UI colour design. This could be due to inadequacies in either the product or the empirical design. The failure led the research to review the previous experiments before deciding on the general design features of the current study. In the first two experiments, an attempt was made to determine colour effect against other conditions such as monochrome. While Experiment 1 was a success in demonstrating the effect of colour against monochrome, Experiment 2 failed to replicate this. As a result, the aim of Experiment 3 was to find an effect of colour. Colour being the focus of this research, it was aimed at producing an effect as big as possible between two types of colour conditions (see next section) in an application environment using a simple colour scheme.

The design rationale is explained in section 8.2.1. Since the generality of the UI colour design failed, the experiment aims to at test the sensitivity of a product design conforming to a single colour rule - **a minimal product**. In this case, the experimental design drives the product design. This is in contrast to the UI colour product used in previous studies where the delivery requirements drive the simplicity of the product design.

The failure of the proposed novel use of colours prompted the research to further review the potential uses of colour demonstrated in HCI research. Implicit to the UI colour design is the use of colour in grouping. Experiment 3 attempts to establish at least the basic crude grouping effect using the minimal product. It was thought that, if the study failed to demonstrate any effect then it would not be feasible to exploit colour grouping for novel functional uses of colour in this research. The thesis would then need to explore reasons for not obtaining the colour effect contrary to the already established results on colour [Treisman 1982].

8.2 Design

In Experiment 3 users were expected to identify a given word - *the target word* - from a collection of words on the computer screen. This experiment examines the test measures between two conditions: 1) A consistent target colour condition in which the test user associates a target word with the target colour feature 'yellow' and 2) An inconsistent target colour condition in which the target word is randomly associated with any of the three different colours - light blue, pink and light yellow (see section 8.2.3)[1]. The target words are the labels of selectable buttons, which are distributed across the computer screen in different colours. The screen layout (see Figure 8.1) is described in detail in section 8.2.3.1. In the first condition, the target colour was made explicit to the test users and they were informed about associating target words with the target colour. This condition was a strong test for the crude grouping effect of colour (see Figure 8.2). In the second condition, test users were not told that the experiment was about colour and simply asked to search and identify target words, one at a time. As described in the next section, the experiment used ten trials each containing one target word. The target word differed from trial to trial. In each trial the location of the target word was changed at random in order to over-ride any preferential location of the target word. In each of the trials, the test user was therefore required to perform a new search for the right target word.

Hypothesis:

Experiment 3 hypothesised that: 1) there would be a colour effect, 2) the users can learn and apply a single colour rule and 3) the test users in the consistent condition would perform better than those in the inconsistent condition. The experiment was intended to produce large differences between the two groups.

The study made use of a task specifically designed for testing the crude grouping effect. It did not purport to be a representative of any real world applications and therefore was not concerned with the realism of the application.

---

[1]For the sake of convenience, the thesis refers to them as blue, pink and yellow.

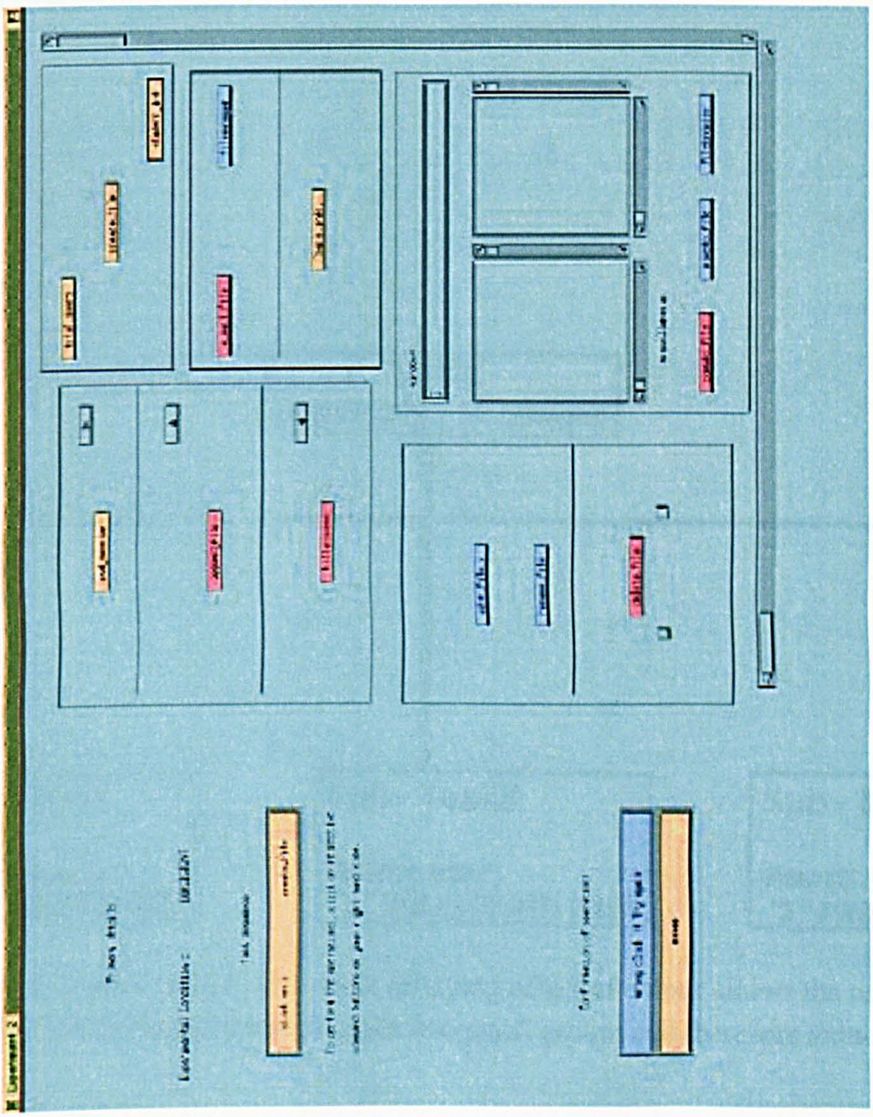


Figure 8.1: Screen layout of the test task in Experiment 3: The screen was divided into two main parts: 1) The left part of the screen contains instructions and a sequence of operations to be performed by the users; and 2) the right part of the screen comprises the searchable area which allowed the users to search for the target word.

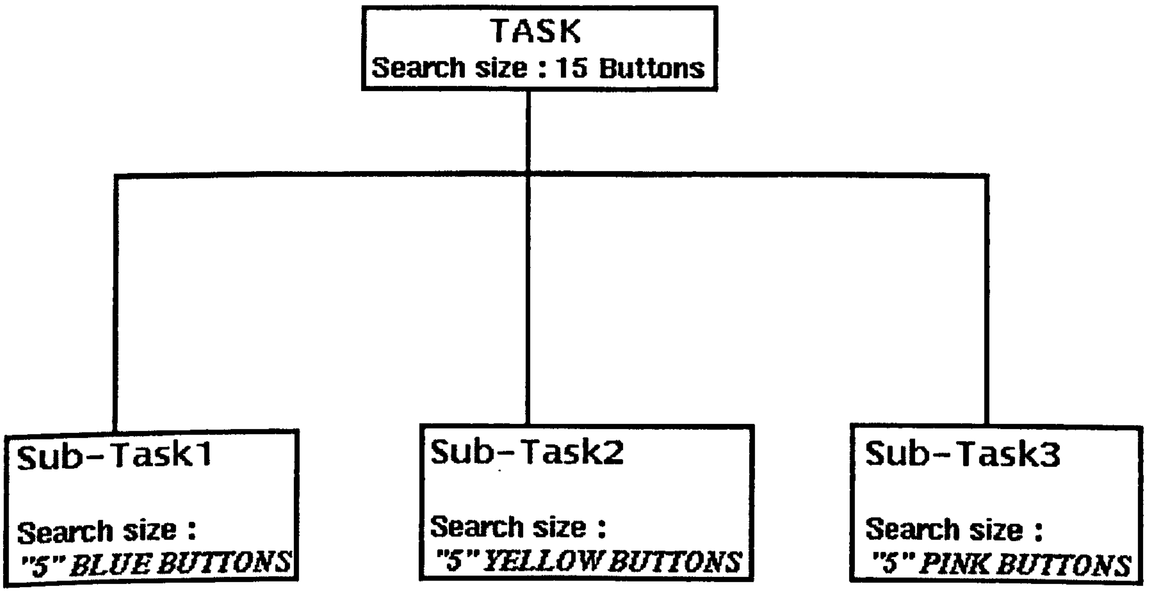


Figure 8.2: User's model: The crude grouping effect of colour allows the user to decompose and organise the total search space into small groups and therefore reduces the search time.

8.2.1 Design Rationale

The primary reason for designing a minimal product is to evaluate users' ability to learn, apply and remember a single colour rule that comprises mapping between a simple perceptual category (yellow) and a simple cognitive category (yellow is selectable). That is, the colour rule specifies that only yellow is selectable and all other colours are non-selectable. The test is therefore conducted using a simulated application environment so as to obtain the maximum crude grouping effect.

The results of this study, if successful, will form the basis for testing other novel, potential uses of colour described in Chapter 4.

8.2.2 Method

The test users were randomly assigned to one of the two groups, each representing a test condition described in the beginning of section 8.2. The group representing the consistent target colour condition served as the 'control' group and the other as the 'experimental' group. Both the groups used the same machine with identical colour displays.

On any one test condition, the user was required to perform a single task. The task consisted of ten trials. The user had to identify a unique target word in each trial. In the first condition, the user applied the single colour rule to reduce the search for the target word. In the second condition, the user had to search all the words to locate the target word in each trial.

8.2.3 Materials

The study was similar to the previous experiments in preparing handouts and using the computer equipment except that no video camera and television were used for data recording and analysis. Instead, the study used automatic logging procedures (written in C and integrated with UIMX software) to collect different performance measures such as task completion times. From the beginning to the end, details of users' selections (such as clicking on a specific button) along with the time details were logged. From the log, it was possible to determine the buttons selected by the user, whether it was right or wrong selection, the time of clicking etc.

The following subsections describe the visual details of the computer screens used in performing the task.

8.2.3.1 Design of the Test Task: Ten Trials

Task Interface:

The eighteen button design, with blue, pink and yellow as background colours, was the default screen design (see Figure 8.1). This was developed using UIMX. Each trial used the default screen design with the following set of design parameters:

- **Target feature:** Background colour.
- **Number of colours:** Three background colours Blue (B), Pink (P) and Yellow (Y) were used in each trial. In the consistent colour condition, users were expected to learn and apply the single colour rule: the target word was always on a yellow background.
- **Graphic design:** Buttons (Motif). All the ten trial designs included identical screen layouts with target words spread randomly in different locations so as to over-ride any location biases. For example, users with a top-bottom bias tend to scan the items row by row from top to bottom. If the target word is always located in the top rows it would bias the results.
- **Number of target words and target colours:** There were ten unique target words, one in each trial. The labels were chosen to represent some familiar operations to most computer users, for example file operations: create file, append file. In the consistent condition, the colour of the target word was always yellow. In the inconsistent condition, there was no unique target colour.
- **Number of options:** In each trial, test users were asked to make a selection among 15 buttons. These buttons were organised into three groups of five buttons each in Blue, Pink and Yellow colours.

The screen was divided into two main parts: (1) The left part of the screen contained instructions and a sequence of operations to be performed by the users; and (2) the right part of the screen comprised the searchable area which allowed the users to search for the target word. The left part included three buttons of equal size. They were labelled with specific instructions on selecting the target word. The right part consisted of fifteen buttons also of equal size. The buttons were labelled with various words, one of them being the specific target word.

Consistent Target Colour Condition:

In this condition (see Figure 8.1), the left side of the screen always displayed three buttons: two of them were yellow and the third in blue. The blue button is non-selectable and reports the status of users' actions in each trial. For example, upon a successful selection of the target word, the blue button would display a confirmation and instruct the user to proceed to the next trial by clicking the yellow buttons. In each trial, the buttons on the right side of the screen were coloured randomly so as to obtain five yellow (target colour), five blue and five pink buttons. The test users were instructed to (i) remember that yellow was the target colour i.e. the target would always be on a yellow button and (ii) search for the current target word among the yellow coloured buttons on the right side of the screen.

Inconsistent Target Colour Condition:

This condition replicated the design of the consistent condition except that in each trial, the target word appeared on any one of the coloured buttons thereby controlling for any location bias. Also, in each trial, the test users were required to perform a new search to find the current target word. In the worst case, the user may end up examining all 14 buttons prior to discovering the 15th button labelled with the required target word. It should be noted, that since the test users were not informed that the experiment was about colour, learning and consequently task performances were not expected to be as good as in the first condition.

Performance of the Task:

On any one particular condition, the test user was required to search and select the correct target word among the fifteen buttons located on the right part of the screen. In each trial, a different target word was given and therefore the users were required to perform a new search. The grouping effect of colour was expected to reduce the search space to five buttons by allowing the users to organise the total search space (i.e. fifteen buttons) into three groups (i.e. each containing five buttons) as shown in Figure 8.2. Users in the consistent target colour condition were expected to remember and apply the colour rule while trying to identify the target word (see Figure 8.3). Analysis of search patterns adopted by users (horizontal or vertical) were beyond the scope of this experiment.

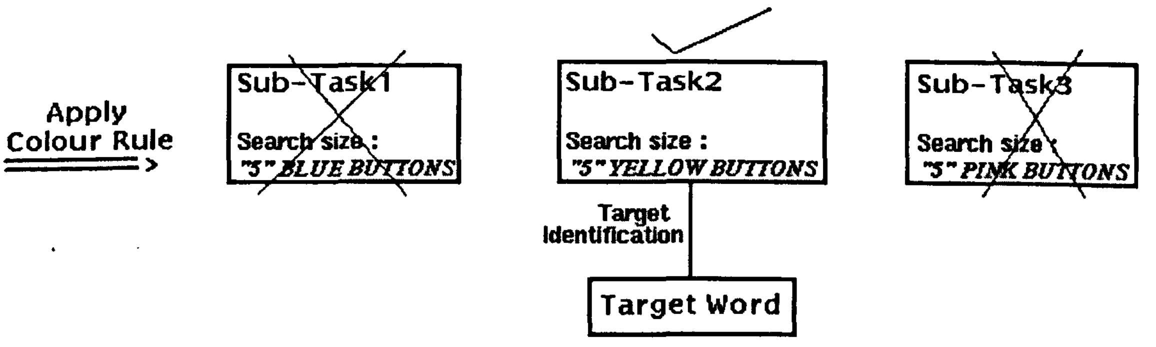


Figure 8.3: The colour rule facilitates the perceptual processes of searching for a target word and thereby makes the search easy by reducing the total search space.

8.2.4 Testing Procedure

The experiment consisted of three phases as described in Table 8.1. The first phase was similar to that of Experiment 1 (see sections 6.2.4). In the second phase, the users were trained on the test task. The training started with demonstrating to each user on how to run the computer program used in conducting the study. After giving a few instructions about the experimental task and prompting the user to input necessary information, such as typing their name in the text input area, the training task was introduced to the user. Users were trained to run through six trials which were similar to the trials in the testing session. Users in the consistent target colour condition were instructed to apply the yellow colour rule to search for the target word.

The third phase was the testing phase. The users were instructed to execute the test task running on the same lines as the training task. They were asked to complete the task as fast as possible while avoiding errors (i.e. non-optimal responses). The computer automatically logged the user's performance.

The above procedure was conducted with one user at a time. At the end, data logs were analysed to obtain task performance measures. The results obtained and conclusions drawn from the statistical analysis of data are reported in sections 8.3 and 8.4.

Experimental Phase	*Duration (Minutes)*	*Description*
1	10	Introduction to the study and the test task
2	10	Training users on the test task
3	20	Testing users on the test task

Table 8.1: Experiment 3 - Description of phases.

8.2.5 Experimental Design

The experiment was carried out under the two test conditions mentioned earlier. The log data from the users was analysed to obtain total task completion times and non-optimal counts. This data was subjected to statistical analysis to measure performance differences in both the test conditions.

8.2.6 Test Users

The experiment was conducted on twelve test users aged between 19 and 25; five were female. All the test users had normal colour vision (see section 5.1.3.3) and none had participated in the prior experiments testing colour. They were randomly assigned to the two experimental conditions, with five test users in each. The IT experience of the users varied from one year to four years.

8.3 Results

Task completion times and non-optimal counts were analysed in this study. The Mann-Whitney U test (see section 6.2.7) was used to determine the difference between the performance measures in both the test conditions. The test was one-tailed because the users in the first condition were predicted to perform better than those in the second.

8.3.1 Task Completion Times

The primary measure was the time taken to complete the test task in seconds. The first and second conditions took 87.2 and 119 seconds respectively as shown in Figure 8.4.

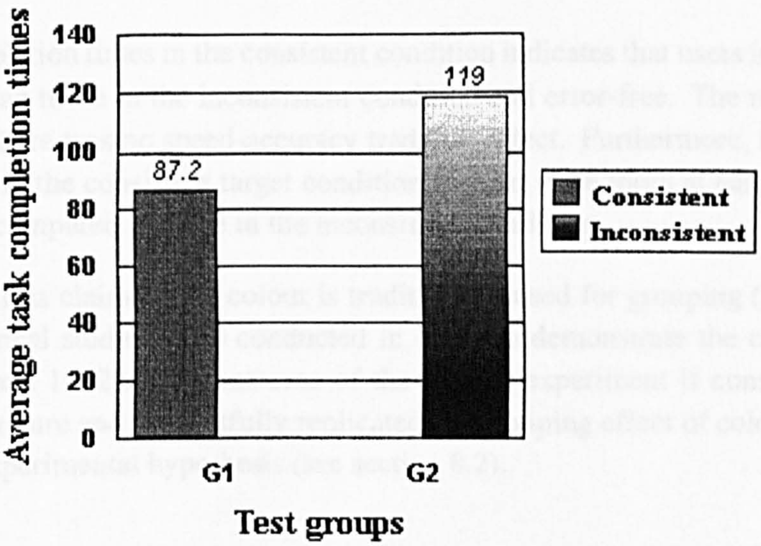


Figure 8.4: Test conditions and average task completion times in Experiment 3.

The U test indicates a significant difference (the Mann-Whitney U value of 4 and a one-tailed probability of 0.013) between the task completion times between the two groups.

8.3.2 Non-optimal Responses

A non-optimal response was logged whenever a user attempted to the select a wrong target word. There was only one non-optimal response in the consistent target colour condition and none in the other condition.

8.4 Conclusions

With the minimal product, the study predicted a crude grouping effect from the work of Treisman & Gelade [Treisman 1982] in terms of a significant difference between users' task performance times of the test conditions. The results of the experiment validated and confirmed the predictions. The absence of colour grouping in the inconsistent target condition affected the performance of test users resulting in high task completion times. In addition, the experiment demonstrated that the users were sensitive to a simple colour rule.

The low completion times in the consistent condition indicates that users in this condition were faster than those in the inconsistent condition and error-free. The results therefore indicate that there was no speed-accuracy trade-off effect. Furthermore, it was observed that the users in the consistent target condition seemed to be more at ease in performing the test tasks compared to those in the inconsistent condition.

In summary: It is claimed that colour is traditionally used for grouping (see Chapter 2). Several empirical studies were conducted in order to demonstrate the colour grouping effect [Treisman 1982]. The outcome of the present experiment is consistent with the published literature and successfully replicated the grouping effect of colour. The results support the experimental hypothesis (see section 8.2).

8.5 Empirical Evaluation of UI Colour Design in Retrospect

One purpose of this research is to exploit novel, functional uses of colour and develop products i.e. appropriate designs that could be delivered to users. The UI colour design is intended to be a pilot product. An important requirement of product delivery is to establish its validity through a careful empirical approach. Validation determines the generality of the product. The first two experiments reported in Chapters 6 and 7 evaluate the UI colour design in this direction.

Experiments 1 and 2 examined the generality of the colour design under multiple test conditions. The results however failed to demonstrate that the design can be simply applied to any general application. In other words, the product can not be delivered to users as capable of application to any user interface. Although, the proposed novel use of colour did not serve the intended purpose, Experiment 3 succeeded in demonstrating the already established use of colour in grouping.

Overall, the research does not recommend the delivery of UI colour design due to its lack of generality. The traditional usage of colour in crude grouping, however, could still be exploited on other platforms such as the Web (see Chapter 4).

Chapter 9

Web Experiment 4: Single Colour Rule

9.1 Empirical Approach: Evaluation of the Web Colour Product

The high level design details of the main product of this research, A Web colour scheme, are described in Chapters 4 and 1. The issues discussed on delivery, design and the empirical approach in the context of the pilot UI colour design (see section 6.1) can be extended to the Web colour design. The empirical approach taken to validate the UI product forms the basis for the current series of usability studies.

One important aspect of the Web colour design is that the perceptual category (i.e. colour) is mapped on to a complex function (i.e. type of a Web link, see section 4.5). This is in contrast to the pilot UI design where the mappings were simple (i.e. one to one) between the perceptual category 'colour' and the function 'input type'. The failure of the generality of the UI design (see Chapter 7) has raised some serious concerns about the delivery of Web design; the reason being that colour is a semantically overloaded feature in the Web design. For example, several types of resources (such as an under construction page or a data dump) could be associated with a blue colour link. Similarly, a yellow link may indicate either a small text-alone document or an important (i.e. a high desirability) graphics image readily accessible on a local site. The complexity of these mappings and their influence on users' task performance require careful evaluation. The empirical studies described in the current chapter and Chapters 10, 11 and 12 provide a rich, evaluation of the Web product and the proposed means of delivery.

9.2 Foreword: Experiment 4

This chapter describes the first usability study in the series of Web experiments. The study replicated the general design features of Experiment 3 (see Chapter 8) on the Web platform. Experiment 3 had successfully demonstrated the colour grouping effect using a minimal product, the single yellow colour rule. This experiment aims at producing the same effect with a similar minimal product on the Web, in a simple task environment. This Web minimal product could be described as a partial implementation of the main Web colour product described in Chapter 4. Although, the Web pages used in this experiment included all the three colours - blue, green and yellow - only a single colour rule was implemented. The rule specifies that only yellow colour links are selectable. In contrast to the mappings described in Chapter 4, the mapping in the minimal product was kept simple i.e. categorical to categorical, between the yellow colour and the associated function (indicating link type i.e. whether a link is selectable or not). Several Web pages were designed for the experiment using the yellow colour rule. They were designed to be realistic and included usual Web components such as text, graphics and hyperlinks.

Overall, the study comprised two parts involving the collection of data. The first part served as a pilot study. Data collected in this part was analysed in order to find differences between the performance measures (see section 9.4.1). Following this, the second part was carried out with more test users. Both the parts were identical in general design, experimental method, materials and test tasks. As in Experiment 1, data collected from both the parts were combined, and analysed to determine the colour effect between the two test conditions, explicit and implicit (see section 9.3). This study forms an appropriate basis for designing the subsequent tests for evaluating the Web colour product (see Chapters 10, 11 and 12).

9.3 Design

The study examines user performances between two test conditions: (1) Explicit colour (2) Implicit colour. The users in both the conditions used identical displays and carried out the same tasks. In the first condition, the colour rule details were made available to the test users. The rule was simple - only yellow coloured (background or border colour) links are selectable. In the implicit condition, the colour rule details were not given to the test users.

Hypothesis:

The study was carried out with the hypothesis that: (1) the experimental design will detect differences in colour in terms of test measures such as task completion times; (2) users can learn and apply the yellow colour rule; and (3) users in the explicit condition will perform better than those in the implicit condition.

The study made use of the curriculum vitae of a Web user in designing the test task. The task was chosen to be representative of the typical use of the Web such as browsing through different types of links and filling forms. The task included answering five simple questions while browsing through the curriculum vitae using the Netscape 3.01 Web browser. All the questions were of a single fact retrieval type (see section 9.3.4.4 for details). They required the test users to locate the correct answer embedded in the text displayed on the screen. The Web pages were designed to require two types of user input: text input which required users to use a keyboard, and selection of hyperlinks using pointing and selection devices, such as a mouse.

9.3.1 Design Rationale

The rationale described in Experiment 3 for testing the effect of colour using a simple, minimal product in a simple environment was applied to the present study. The main reason for this initial design simplicity was the failure of Experiment 2, which involved testing the colour rules under multiple test conditions in an environment which was relatively complex. The subsequent research intends to test a simple colour product before testing the main product which involves complex stimulus-perceptual mappings. If the simple colour associations fail, then the experiments aiming at testing the complex colour associations should not be conducted.

The purpose of the study was twofold: Firstly, the study aimed at carrying out an initial testing of the sensitivity of the colour metric with respect to users' task performance measures. Once the sensitivity was established, the study intended to find the colour effect on users' task performance and their ability to learn/infer the rules while performing the task. Learning is one of the delivery requirements that drives the product as well as the experimental design. The study comprised the two test conditions mentioned previously in order to determine whether the colour rules need to be taught explicitly to the users or they can infer the rules themselves when colours are used consistently across the pages.

Some reasons for using the curriculum vitae in designing the test task include: (1) Generality of the colour rules - the study intends to use some real Web pages in order to demonstrate the generality of applying the colour product to any Web page, keeping the original contents intact and; (2) The ability to use graphics to represent hyperlinks - the Web pages included graphic buttons (Java applets) and image maps in GIF format. The proposed colour associations could be easily implemented by editing background colour values of these graphic forms or placing a border around them.

9.3.2 Method

The test users were randomly assigned to one of the two test conditions. The group of users representing the explicit test condition served as the 'control' group and the other the 'experimental group'. Both the groups used the same machine with identical colour displays (see section 5.1.3.3).

On any particular condition, the test users were required to perform an identical test task. The task consisted of eight different pages; five of them comprised questions. Each page contained multiple links and the user had to select the target yellow link to proceed further. In the first condition, the user applied the yellow rule. In the second, the selection of the target link was not straightforward; the user had to select the link either randomly or based on individual preferences or his/her understanding of the features used in the display.

Both the groups were required to perform the task on identical colour displays. A performance difference was expected due to the difference between explicit and implicit conditions. The explicit group users were, of course, expected to perform better relative to the implicit group users. The latter was informed that this experiment was about using colour in Web pages, but were not given the other details. The possible implications could be that some users might have considered use of colour from an aesthetic standpoint while the others might have made an attempt to see its usage beyond aesthetics.

9.3.3 Materials

The test users performed the given task in a computer laboratory using the Web on a HP Unix colour workstation (see section 5.1.3.3). Handouts consisting of the relevant details of the experiment were first given to the users. The same computer was used throughout the study and the machine was equipped with a keyboard and a mouse. Netscape Web browser 3.01 (the version available at the time of conducting the studies) was used to browse through the Web pages and user performance was automatically logged using CGI scripts running on the local Web server. Data sheets were used to record users' comments.

The following subsections describe the visual details of the Web pages designed for the experiment.

9.3.3.1 Design of the Test Task

Design of the Web Pages for the Task

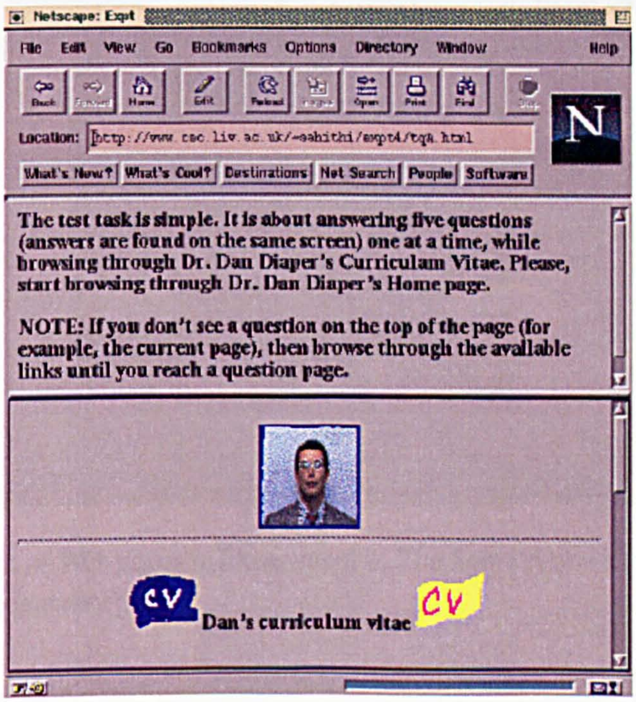


Figure 9.1: Design of Web pages in Experiment 4: The figure shows a scaled down screen dump of an example non-query page.

The Web pages (see Figures 9.1, 9.2 and 9.4) consisted of a number of links represented by simple graphic designs, for example buttons (Java applets) and small image maps, painted

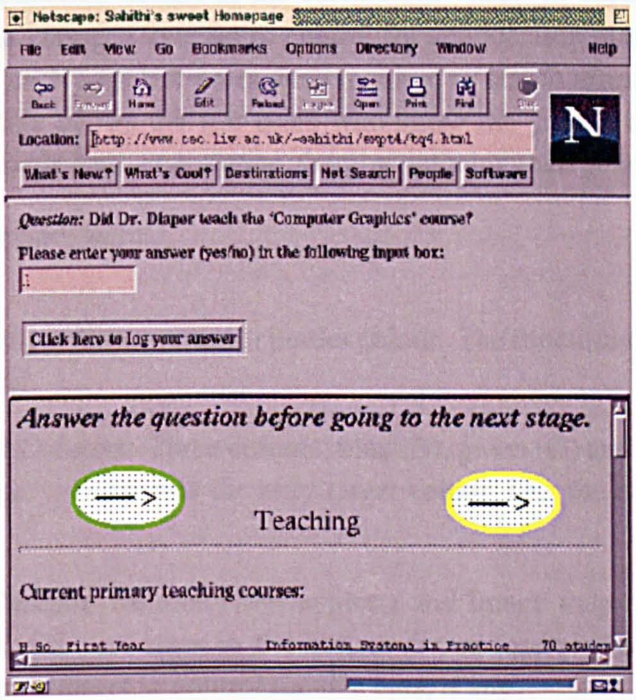


Figure 9.2: Design of Web pages in Experiment 4: The figure shows a scaled down screen dump of an example query page.

in three different colours: blue, yellow and green (see section 5.1.3.3). The target colour was always yellow and the target features were background and border colours. Each page included only one selectable link and it was always painted in a yellow background or border colour. The study used five query pages and several non-query pages each of which contained only one target link. In each page, the target link was labelled appropriately so as to represent the contents of the resource it was pointing to. For example, personal information or education details in the curriculum vitae. In each page, the location of the target link was changed in order to control for any location biases (see section 8.2.3). Consequently, in each page test users were required to search afresh for the correct target link. However, this allows the test users who know the yellow colour rule, to quickly identify the selectable link, *at a glance*, thereby reducing their search time.

The design parameters were:

- **Target feature:** Background or border colour. The functionality of both the features was identical.
- **Number of Colours:** Three colours, blue (B), green (G) and yellow (Y), were used in each page. yellow was the only target colour, i.e. the colour of the selectable links.
- **Graphic Design:** Buttons (Java applets) and image maps. Web pages included graphics of different sizes in the bottom frame (see section 9.3.3.2 for the page design) in an attempt to control for size bias, if any.
- **Number of Options:** In each page, users were asked to make a selection from available options. The number of options on any page varied from two to ten. Most of the pages had less than five.
- **Number of Target Links:** In each page, test users were presented with a single target link, which was always yellow.

Each query page (see Figure 9.2) had two or three colours, and so, two or three links. Query pages 1 and 4 had blue, yellow links; 2 and 5 had green, yellow; and 3 had all the three colours. The target link (yellow colour) was placed randomly in all query (see Figure 9.2) and non query pages (see Figure 9.1). In query pages it was located in the middle in pages 1 and 3; on the left in 2 and 5; and on the right in 4.

9.3.3.2 Sequence of Operations in Performing the Task

The design of the task consisted of three phases. In the first phase, the Web pages contained general details of the study and some typical Web features (such as hyperlinks represented in different forms - buttons, icons and images, etc.). The aim was to familiarise the test users with the relevant features of the experiment and make sure that they acquire basic skills of selecting different forms of links while browsing through the Web pages.

In the second phase, the users were trained on six Web pages consisting of sample questions and related text. The training pages also included both query and non-query pages. In the former, each page consisted of two frames. The top frame contained the question to be answered by the user and the bottom one included graphic-links and some text containing the answer. Once the user found the correct answer, he/she was directed to type in the answer in the text field of the top frame and confirm the selection by clicking on the log button. Details such as the answer and the time of entry were automatically logged for each user. The non-query pages comprised some text and graphic buttons representing hyperlinks. The users were required to make a selection among the displayed links and continue browsing until they reach a query page. All the query pages used image maps representing hyperlinks. A test user was required to identify and select the clickable region of the image map in order to proceed to the next page. Whenever a user failed to make a correct selection, a new screen was displayed with a message indicating the wrong selection and the user was prompted to backtrack to the previous query page (see Figure 9.3). Test users in both the groups were trained on using these pages and answering questions. The users in the explicit condition were given specific instructions on the target colour feature as described later in this section.

In the third and final phase, the users were tested on the Web pages. The test pages were similar to the pages in training section so that the users were guided through a series of five questions with minimum assistance. Initially, the users were presented with non-query pages containing five or more graphic links accompanied by some text. They were required to select the yellow coloured button or image map and continue browsing until they hit upon the next query page.

In summary, each user performed the following sequence of operations during the testing (see Figure 9.3):

- Browse through the pages until a query page was reached.
- Read the question.
- Go through the textual information to find the answer.

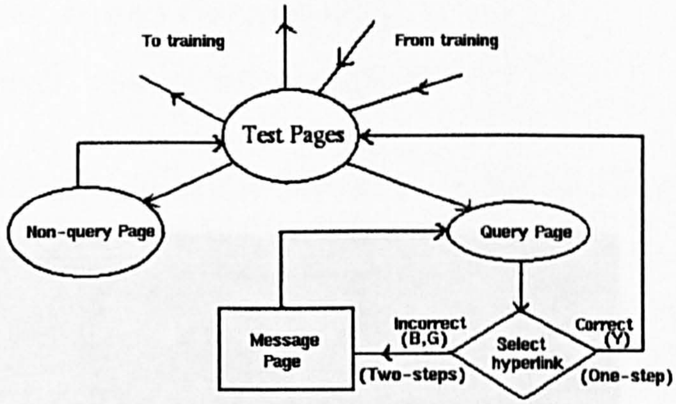


Figure 9.3: Design of Web pages in testing phase. Test users encounter both query and non-query pages and were required to select target links (yellow) to proceed to the next page.

- Type in the answer in the text field.
- Click on the log button to record the answer.
- Select the correct hyperlink to proceed to the next page.

9.3.3.3 Colour Learning: Design of Training Task for the Explicit Condition

Prior to training the users on performing the experimental task, The test users in the explicit condition were instructed to learn and apply the yellow colour rule. This training was given on five Web pages that enabled the users to browse through the pages by selecting links painted in the target colour, that is, yellow background or border. Test users who were quick in learning the yellow colour rule were provided with an option to discontinue the learning session and go on to the testing phase. Test users in the implicit condition were excluded from this colour learning session.

9.3.3.4 Type and Difficulty of Questions in the Task

All the five questions were of the 'simple fact retrieval' [Spool 1997] type for which there was only one correct answer. Typical questions were:

- Where was Prof. Diaper working?

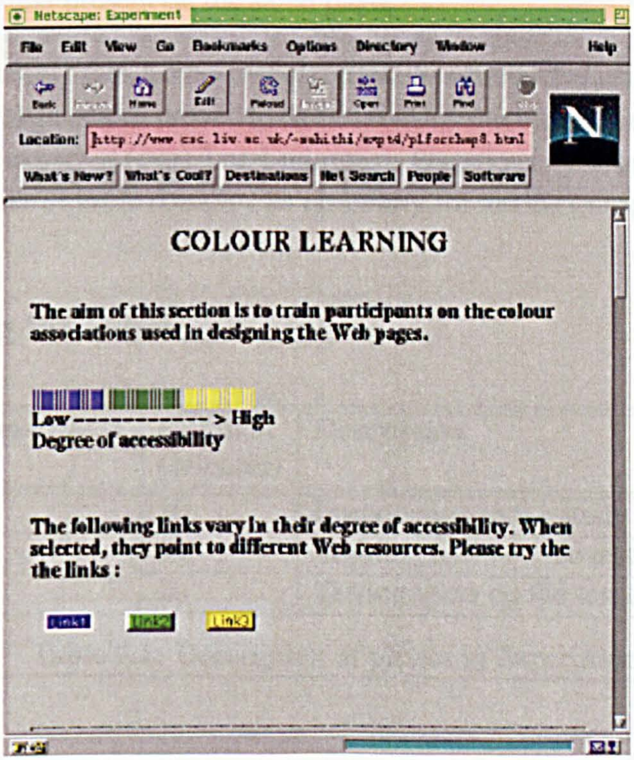


Figure 9.4: Design of Web pages in Experiment 4: The figure shows a scaled down screen dump of an example page used in the training session in order to teach colour rules to the users in the explicit condition.

- What was Prof. Diaper's PhD area? and
- Where did Prof. Diaper work during 1984-86?

Each question was designed to be reasonably simple so that a user could answer the query after reading the text present on the same page.

Before testing on real users, a rehearsal of the test was conducted using some Web-veterans. Many of the model questions proved to be easier than anticipated. Some questions however took longer than expected (over 3 or 4 minutes). Questions were modified or discarded to make them simpler or more complex as necessary for testing purposes.

9.3.4 Testing Procedure

Experimental Phase	*Duration (Minutes)*	*Description*
1	10	Introduction to the study and the test task
2	10	Training users on the test task
3	20	Testing users on the test task

Table 9.1: Description of phases in Experiment 4.

This study consisted of three phases as described in Table 9.1. The first phase started with the simple test for normal colour vision. Next, the handouts comprising details of the study were distributed. The user was then introduced to the study and the test task.

The second phase consisted of two different training sessions, one for each group. The session for the explicit group included training users on both the task and colour rules (see previous subsection). On the other hand, the session for the implicit group did not include colour training.

The training and testing was carried out on a local Website. It was made clear to the users that the answers were always available somewhere within the same Web page that contained the question. The experimenter did not explain the different testing conditions to any of the test groups. Only the users in the explicit condition knew that the experiment was about the functional use of colour. In addition, users were notified about the three prerequisites of the study: 1) the top menu bar in the Netscape browser must not be used while performing the tasks, 2) the transfer of data must not be interrupted while Netscape is downloading the contents and 3) test users need to complete all phases of the experiment.

Users were run through the three phases in sequence. Initially, the experimenter demonstrated to each test user how to use the machine and run the Netscape browser to access the experimental Web pages. After giving a few instructions about the experimental task and prompting users to input necessary information, the session introduced the training-task in order to familiarise all users with the experiment before any performance measurements were made. Later they were asked to go through the test-task. Since the research was not specifically testing Web browsing capabilities, the experimenter assisted those users who had any browser related difficulties, for example the use of scroll bars, page-up and page-down keys. This was allowed so that users would not associate a page design with a problem that actually pertained to the Web browser.

While the testing was in progress, interaction between the experimenter and the users was minimised. Users were asked to work through the test tasks as much on their own as possible. Where available, the study recorded the comments made by the users.

Each user went through all the Web pages in the time allotted. All of them followed the same order in which the Web pages were tested. Answering a question followed the selection of the yellow coloured hyperlink in one step (on selecting the correct link) or two steps (on selecting a wrong link) to move to the next query page (see Figure 9.3). Non-optimal responses, such as selecting a wrong link, were found to be common among the users in the implicit condition.

At the end of the session, users were asked to type in their names in the text field and log-off from the system. The test data was automatically logged by the system.

9.3.5 Experimental Design

As described in section 9.3.1, the experiment was conducted in two parts the first of which formed the pilot study to determine the sensitivity of the test measures, i.e. task completion times. The pilot study was conducted on eight users and their responses were logged on the computer. The results of this study, along with conclusions, are summarised in Appendix G.

Having established the sensitivity of the test measures, the second part was carried out with four users. This was an extension of the pilot study and the combined responses of all users in both the parts (twelve in all) was subjected to analysis.

Users were asked to identify colours on the screen, as a test of colour deficiency (see Chapter 5). No standard procedures were used as there was no red-green combination.

9.3.6 Test Users

The test was conducted on twelve users; nine male and three female aged between 18 and 26. None had previous knowledge of the proposed use of colour on the Web and none had participated in prior experiments testing colour. Test users were randomly assigned to one of the two experimental conditions with six users in each. All had normal or corrected to normal vision.

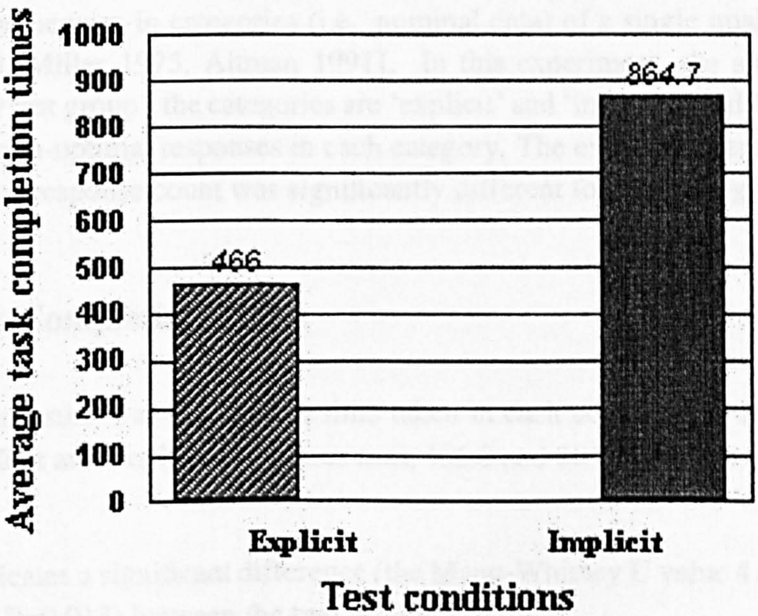


Figure 9.5: Experimental groups and average task completion times in Experiment 4.

Since the research was testing aspects such as colour learning, retention of a single colour rule and the effect of the colour grouping on users performance, users in this study were not required to have any particular skills or any great Web proficiency. The only requirement was to recruit people who were familiar with using a Web browser. The amount of the

Web experience ranged from three months to two years, from less than one hour per week to more than five hours per day.

9.4 Results

The experiment collected data on two key performance measures: task completion times and non-optimal counts. These measures are shown in Figures 9.5 and 9.6. A Mann-Whitney U test (see section 6.2.7) was used to determine whether there was a significant difference between the time measures in both the test conditions. The test was one-tailed because the users in the explicit condition were predicted to perform better than those in the implicit.

The study employs Chi-Square test, an appropriate statistical test that utilises data in the form of frequencies in categories (i.e. nominal data) of a single qualitative variable [Daniels 1979, Miller 1975, Altman 1991]. In this experiment, the single qualitative variable is 'the test group', the categories are 'explicit' and 'implicit', and the data consists of number of non-optimal responses in each category. The experiment examines whether the non-optimal response count was significantly different for different groups.

9.4.1 Task Completion Times

The primary measure was the average time taken in each condition to complete the test task. The explicit and implicit conditions took 466.0 and 864.7 seconds respectively (see Figure 9.5).

The U test indicates a significant difference (the Mann-Whitney U value 4 and a one-tailed probability of P=0.013) between the two groups.

9.4.2 Non-optimal Responses

A non-optimal response was logged whenever a user attempted to select a non-yellow target link. The average non-optimal response counts for explicit and implicit conditions were 0.5 and 5.8 respectively. The Chi square value 4.46 (probability less than 0.05) indicates a significant difference in this performance measure between both the conditions.

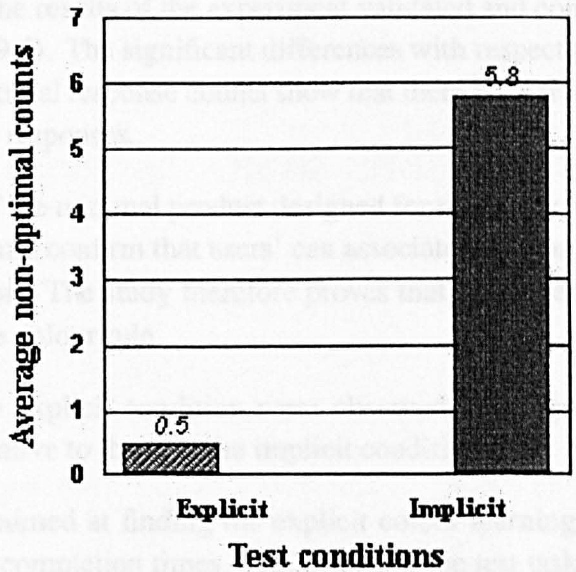


Figure 9.6: Experimental groups and average non-optimal response counts in Experiment 4.

9.4.3 Debriefing

After the test, all users' were debriefed. They were asked for any comments they might have about the study and suggestions for further improvement. Eleven test users appreciated the idea of using colour in a manner that improves users' Web browsing performance. Only one user did not show any interest in learning and remembering colour associations. Two users did not like the shades of blue and green, although the other ten users did not comment about colour shades. This was purely based on aesthetics. The thesis acknowledges the difficulties involved in producing aesthetically pleasing displays for all users. As mentioned in Chapter 4, the colour design could be experimented with other colour sets. This is however not within the scope of the current research.

9.5 Conclusions

Overall, the study successfully proved the hypothesis (see section 9.3). In particular, the study predicted that the users in the explicit condition would perform better relative to the implicit group. The results of the experiment validated and confirmed the predictions (see Figures 9.5 and 9.6). The significant differences with respect to both the completion times and the non-optimal response counts show that there is no trade-off of speed against possible non-optimal responses.

Further, the results of the minimal product designed for this study are consistent with that of the Experiment 3 and confirm that users' can associate a colour perceptual category to a simple Web function. The study therefore proves that users' can learn, remember and apply a single, simple colour rule.

The test users in the explicit condition were observed to more confident in browsing through the pages relative to those in the implicit condition.

The study was only aimed at finding the explicit colour learning effect vs. implicit in terms of overall task completion times. The design of the test task was not driven by the implicit learning condition. As a result, the data was not examined from the standpoint of finding implicit learning of the yellow rule in individual trials.

In summary, the study indicates that people are able to use a simple, single colour rules on the Web. There exists an effect of colour and the associated functionality on users' performance and this effect could be described as a 'robust' one. Having tested for an effect of simple colour rule, the research aims at testing the main Web product that includes multiple colour rules and complex mappings between each colour and the function, such as link type, as described in Chapter 4.

Chapter 10

Web Experiment 5: Three Colour Rules

10.1 Foreword

This experiment continues the sequence of evaluating colour designs on the Web. Experiment 4 demonstrated users' ability to learn the single yellow colour rule with a consequent effect on their task performances. Experiment 5 extends the test by using multiple colour rules and measures resulting effects on the task performances under similar test conditions.

The major difference between the previous experiment and this one was the number of colour rules and the nature of colour associations. This experiment fully implements the proposed Web colour product i.e. the three colour scheme (see Chapter 4) to represent hyperlinks pointing to various types of resources available on the Web. The Web product may be simple in the number of rules, but conceptually the colour associations are more complex than those in the previous experiments. Experiment 4 demonstrated that people could associate a simple perceptual category (i.e. colour) with a simple function. The main purpose of Experiment 5 reported in this chapter was to determine whether people can relate a simple colour category (blue, green or yellow) to a complex function (types of hyperlinks).

The following sections describe the overall design, design rationale, results and conclusions drawn from the statistical analysis. Some details of the study, for example the testing procedure and the experimental design, are similar to those in the previous study and therefore are not discussed.

10.2 Design

The experimental design of Experiment 5 was similar to that of Experiment 4. The current experiment was designed to examine whether there were any user performances between explicit and implicit conditions. In the former, the colour coding scheme for Web links was explained to the users who were instructed to work through the experiment in accordance with it. In the implicit design, the colour scheme details were not disclosed to the users.

There are however a few differences with respect to Experiment 4. Previously, the target colour (i.e. the colour of the selectable links) was always yellow. In the present study, both yellow and green were made the target colours. The colour associations also differed to comply with those in the Web colour product (see Chapter 4). Further, artificial time delays were added upon selecting green or blue links. The experiment again used Dr Diaper's curriculum vitae of Experiment 4 but with a different set of questions (see the next section for the design rationale). There were six questions of the simple fact retrieval type. Test users in both explicit and implicit conditions received the same practice and instructions described in the previous experiment.

Hypothesis:

The study was carried out with the hypothesis that: (1) users can learn the three colour rules and associated complex mappings, and apply them across different Web pages; and (2) explicit teaching of the colour rules will have a significant effect on users' performance. Further the study was intended to establish that the proposed Web product can be generally applied to a Web page without affecting the original contents.

10.2.1 Design Rationale

The curriculum vitae used in Experiment 4 could be described as representative of real Web pages. It contained multiple links; some of them point to small amounts of text, some links lead users to large graphic images and some are under construction links. In the previous study, many pages of the CV were not used since the design of the test task was intended to be very simple. Many pages of the CV consisting of different types of links could be well represented using the three colour rules. As a result, the same CV was used in the present study.

The pages used different types of graphic forms, for example image maps and buttons

(Java applets) representing the hyperlinks. Many of them replicated the features of some of the images found on the Web during the Web survey (see Chapter 4 for the details of the survey) conducted as part of the research. Both the target colour features proposed in the Web product i.e. background and/or border were applied in designing the image maps. This is done either by editing background colour values of the graphics or by placing a border around them. The generality of these features/forms and their usage in the test pages have added realism to the display.

The rationale of the explicit and implicit test conditions in Experiment 4 could be extended to the present study. The mode of training users on the three colour rules is one of the important delivery requirements of the product. The previous experiment demonstrated that explicit learning of the colour rule produces better performance than implicit learning. This experiment was aimed at testing for similar evidence in the complex Web colour product.

In many real Web pages, selection of wrong links often result in much delayed access which earned WWW the sobriquet "World Wide Wait". The purpose of adding artificial delays in the test task design was to give test users a representation of natural delays that are normally experienced while accessing different types of resources (see Chapter 4) located either remotely or locally on the Web. Using built-in delays associated with specific colours in a consistent manner, users were penalised on wrong links or rewarded with quicker access on correct ones. The results of this study forms an appropriate basis for testing many other aspects of the main Web product (see Chapters 11 and 12).

10.2.2 Method

The experimental method followed that of Experiment 4. The test users were randomly assigned to one of the two test conditions. The group of users representing the explicit test condition served as the 'control' group and the other the 'experimental group'. Both the groups used the same machine with identical colour displays.

All users performed an identical test task comprising six query pages and several non-query pages. Each page contained multiple links and the user had to select the target links (yellow or green) to proceed further. Users in the explicit conditions were expected to apply the colour rules. The users from the implicit condition were unaware of these rules and were predicted to make random link selections. The explicit group users were therefore expected to perform better relative to the implicit group users (see experimental hypothesis).

10.2.3 Materials

The materials used for the experiment were almost identical to those in Experiment 4. Test users used a HP Unix colour workstation during the experiment (see section 5.1.3.3). The details of the experiment were supplied to the users in handouts. The responses were automatically logged by the machine.

The following subsections describe the visual details of the Web pages designed for the experiment.

10.2.3.1 Design of Web Pages for the Test Task

The overall design of the Web pages was similar to that in Experiment 4 except for the graphic forms (see Appendix H.2; Figure H.2 for some example images) of hyperlinks described in Table 10.1. The design applied some typical graphic representations of hyperlinks used by Web page developers. Image maps differing in shape and size were used along with graphic buttons (Java applets) directly drawn from Experiment 4. Each page had one or two selectable links painted in yellow/green background/border colour. The non-selectable links were always in blue background/border.

Several Web pages were designed both for the training and testing sessions. The training task contained five pages: three of them were query pages and the remaining non-query pages. The test task included 12 pages: six query (see Figure H.3) pages, five non-query pages (see Figure H.4) and one error message page. Each page had one or two target links which were labelled appropriately so as to represent the contents of the resource it was pointing to, such as personal information or education details in the curriculum vitae. In each page, the location of the target link was changed so as to control for any location biases. The implicit group users were therefore expected to perform a new search for the target link in every Web page. Explicit group users who knew colour rules were expected to recognise selectable links, *at a glance*.

The design parameters were:

- **Target feature:** Background or border colour. The functionality of both the features was identical.

- **Number of Colours:** Three colours, blue (B), green (G) and yellow (Y), were used in each page. yellow and green were the target colours, i.e. the colours of the selectable links.
- **Graphic Design:** Buttons (Java applets) and image maps.
- **Number of Options:** In each page, users were asked to make a selection among several options. The number of options on any page varied from two to ten. Most of the pages had less than five.
- **Number of Target Links:** In each page, users were presented with at most two target links in yellow and/or green.

Each query page had two or three colours (i.e. links) as shown in Figure 10.1. Pages 1 and 4 in the sequence had blue and yellow links, 2 had green and blue, 3 had green and yellow, and 5, 6 had links in all the three colours. The target links (yellow or green) were placed at different locations in all the pages. On query pages, it was located in the middle in page 5, on the left in 2 and 6, and on the right in 1, 3 and 4. The golden ratio (see Appendix H.1) was employed in the dimensions of the rectangular image maps in query pages 1 and 5[1].

Design of both training and testing pages involved adding artificial delays to the response times of the differently coloured hyperlinks. For example, on selecting a link with low accessibility and desirability (represented by blue background/border colour) a user would suffer a delay/penalty of 45 seconds. Similarly, a user had to experience a delay of 20 seconds on selecting a link (green background/border) with medium accessibility (a measure of how fast a link is accessed on the Web) and desirability value as illustrated in Figure 10.1. No delays were added to the yellow links which represent high accessibility and desirability value. For experimental purposes, green and yellow links were designed to point to the same resources with the same desirability value. The user can follow either green or yellow links to complete the task successfully. Blue links do not point to the same resources, but display message pages requiring users to go back to the previous screens and follow the correct links. Consequently, blue links are considered least desirable. The added delays make the accessibility values different for the links and consequently affect the combined value.

[1]The purpose was to create aesthetically pleasing graphics. The thesis assumes that the application of Golden Ratio may not influence task performance times.

10.2.3.2 Sequence of Operations in Performing the Task

The operations performed during the test task followed a similar sequence to the previous experiment. The task consisted of three phases. The first was an introductory phase. The second was the training phase. In this phase, users would go through a sample of both query and non-query pages (see Figure 10.1). In the query pages the users were required to make a selection. The explicit group was expected to apply the colour rules. The implicit group users, who were not trained on colour rules, were expected to make the selection based on their understanding of the task. The task was designed to produce artificial delays upon non-optimal selection (blue or green links).

In the third and final phase, the users were tested on the Web pages consisting of six query pages. Initially, the users were presented with non-query pages containing multiple graphic links accompanied by some text. They were required to select yellow or green coloured buttons or image maps and continue browsing until the next query page. The selection criteria for a target link is illustrated in Figure 10.1.

10.2.3.3 Colour Learning: Design of Training Task for the Explicit Condition

As part of the training phase, the users in the explicit condition were trained on learning and applying the three colour rules. The colour training task contained nine pages. Each page comprised three links, one in each colour i.e. blue, green and yellow. The order of the links varied from page to page: B-G-Y, B-Y-G, G-B-Y and so on. Test users were instructed to learn and apply the target colour rule: selectable links were in yellow or green background/border colour. The blue and green links were associated with artificial delays (see section 10.2.3.1). The users who were quick in learning the target colour rule were provided with an option to discontinue the learning session and go on to the testing phase. The implicit condition group was excluded from this colour learning session.

10.2.4 Test Users

There were fourteen test users, five of whom were female, aged between 19 and 25 years. All were new to the colour studies. Nine had normal vision and five had corrected vision. Their Web experience ranged from ten months to three years, from seven hours per week to more than three hours per day.

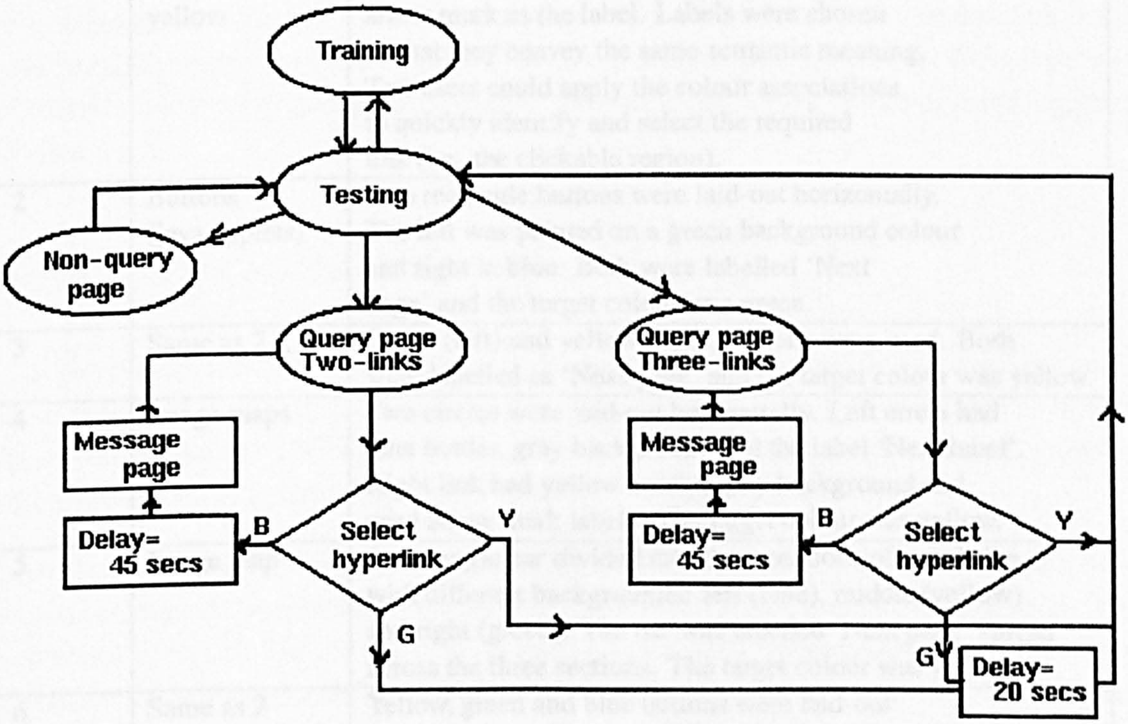


Figure 10.1: Selection criteria for the target link in Experiment 5. Test users encounter both query and non-query pages and were required to select target links (yellow or green) to proceed to the next stage.

Query Page	Graphic Format (Hyperlinks)	Description
1	Image map (Target colour was yellow	A rectangle painted in two background colours: left half in blue and the right in yellow. Left half was labelled as 'Next page' and the right had arrow mark as the label. Labels were chosen so that they convey the same semantic meaning. Test users could apply the colour associations to quickly identify and select the required link (i.e. the clickable region).
2	Buttons (java applets)	Two rectangle buttons were laid-out horizontally. The left was painted on a green background colour and right in blue. Both were labelled 'Next Page' and the target colour was green.
3	Same as 2	Green (left) and yellow (right) buttons were used. Both were labelled as 'Next page' and the target colour was yellow.
4	Image maps	Two circles were laid-out horizontally. Left circle had blue border, gray background and the label 'Next label'. Right link had yellow border, gray background and used arrow mark labels. The target colour was yellow.
5	Image map	A rectangle bar divided into three sections of equal size with different backgrounds: left (blue), middle (yellow) and right (green). The bar was labelled 'Next page' spread across the three sections. The target colour was yellow.
6	Same as 2	Yellow, green and blue buttons were laid-out horizontally. All the buttons were labelled 'Next page' and the target colour was yellow.
Message Page:		This page was displayed to the user when ever a wrong link was selected, for example links painted in blue. The page included a message that directed the user to backtrack to the previous page by selecting the one and only available image map painted in yellow background.

Table 10.1: Representation of hyperlinks in different graphic forms in query and message pages (see Appendix H.2).

10.3 Results

The two test measures used in the analysis were average task completion times and non-optimal response counts (see Figures 10.2 and 10.3). Task completion time, measured in seconds, was defined as the interval between the start and end of all the six questions.

An optimal response was counted when the test user selected the target hyperlink and/or entered required information for each question. On the other hand, a non-optimal response was counted when a wrong hyperlink or a non-selectable region in a Web page was selected[2]. Users' responses are analysed to determine the number of non-optimal response counts.

10.3.1 Task Completion Times

The mean response times for answering all queries for both explicit and implicit conditions were 173.2 and 557.8 seconds respectively. These times included artificial delays described in sections 10.2.1 and 10.2.3.1. The mean completion times, after removing the delays for both the groups were 135.4 and 357.2 seconds respectively.

The Mann-Whitney U test (see section 6.2.7) was applied to find out whether there was any significant difference between the time scores of the two experimental groups. The test was one-tailed. The test results indicated a significant difference (the Mann-Whitney U value of 0 with a probability less than 0.001) between the time scores (both with and without added delays) of the two groups. Figure 10.2 shows average response times of both the groups with and without delays. It took significantly less time for the users in explicit condition to locate the desired link in a page.

Figure 10.4 and 10.5 show response times for each query with and without induced delays. Spearman's rank correlation coefficient is calculated for the explicit and implicit query response times in Figure 10.5. The test indicated that there is no significant correlation (correlation coefficient = 0.772, N = 7) between the response times, even though the graphs appear correlated. Lack of a significant correlation is probably due to the small sample size.

[2]Note: It is not possible to calculate the maximum number of non-optimal response counts because mouse clicks on non-selectable regions in rapid succession were found to be very frequent among the users.

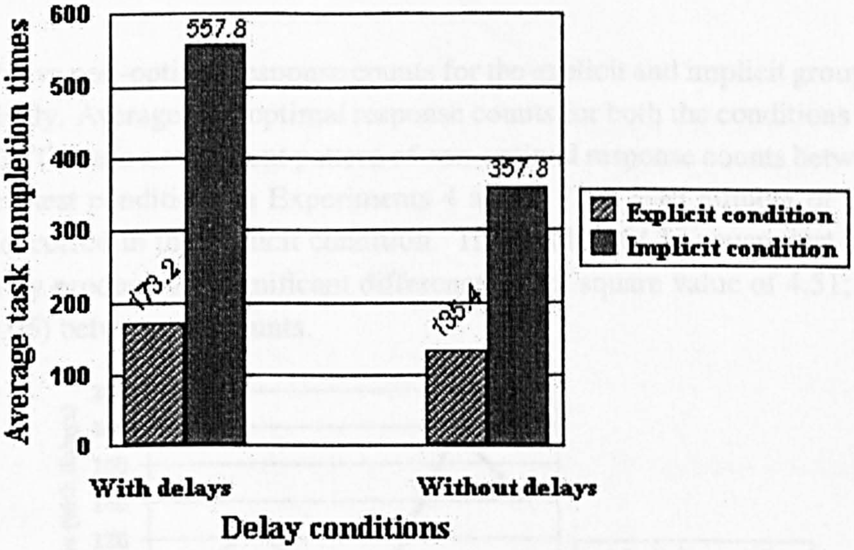


Figure 10.2: Test conditions and average task completion times in Experiment 5.

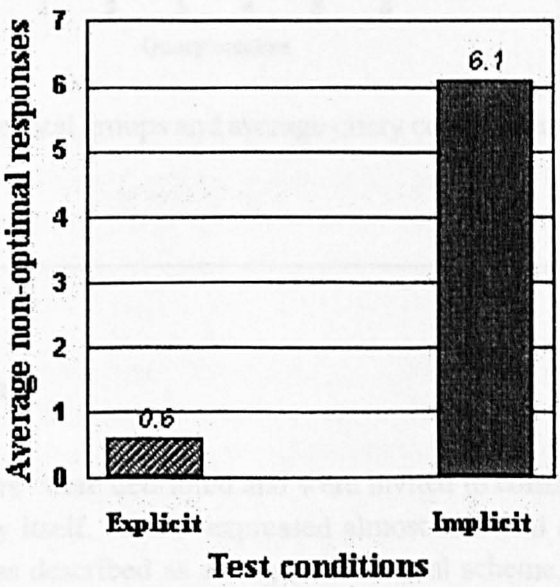


Figure 10.3: Test conditions and average non-optimal response counts in Experiment 5.

10.3.2 Non-optimal Responses

The cumulative non-optimal response counts for the explicit and implicit groups are 9 and 37 respectively. Average non-optimal response counts for both the conditions is shown in Figure 10.3. There is a consistent pattern of non-optimal response counts between explicit and implicit test conditions in Experiments 4 and 5. The least number of non-optimal responses occurred in the explicit condition. The results of Chi-square test supports the hypothesis by producing a significant difference (a chi square value of 4.51; probability less than 0.05) between the counts.

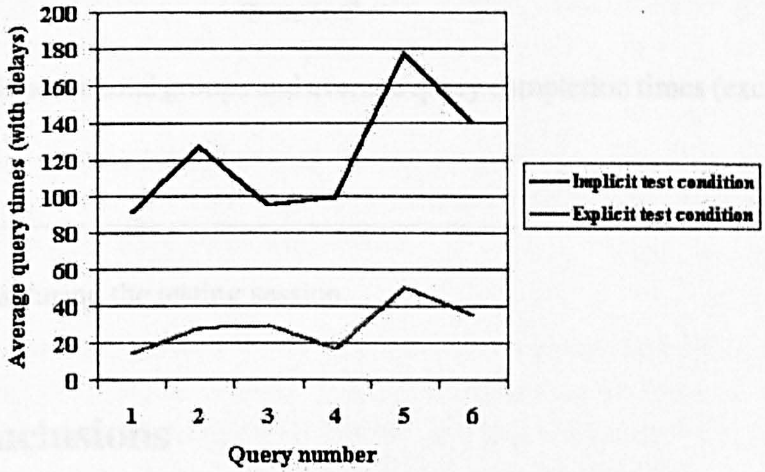


Figure 10.4: Experimental groups and average query completion times (including delays) in Experiment 5.

10.3.3 Debriefing

After the test, all users' were debriefed and were invited to comment on the three colour scheme and the study itself. Users' expressed almost identical opinions about the Web colour scheme. It was described as a simple and useful scheme. Users found it easy to learn and remember the three colour rules. Interestingly, a couple of test users compared the scheme to that of traffic colour signals. They suggested exchanging the proposed uses of green and yellow, thereby making the green the target colour of choice. This preference for green over yellow, however, did not affect the users' learning and the application of

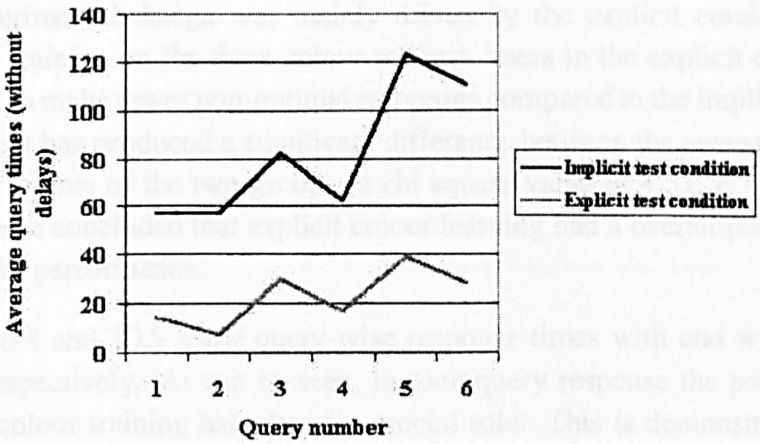


Figure 10.5: Experimental groups and average query completion times (excluding delays) in Experiment 5.

the colour rules during the testing session.

10.4 Conclusions

Overall, the study supported the hypothesis (see section 10.2). The study predicted that the explicit group would perform better than the implicit group. The statistical analysis confirmed the predictions and drew some interesting inferences in addition to those presented in section 9.5.

- Task completion times were determined for the experiment in two ways: (i) with added artificial delays and (ii) without delays. In both the cases, the test produced a significant difference between the time scores. As mentioned earlier in section 10.2.1, a delay is a penalty on selecting a blue or green link. Yellow links do not suffer any such time penalties. Time difference between the total time and added delays determine the actual time spent by a user in search and identification, and decision making. The test shows a significant difference between the time scores (i.e. search and decision times) of the two groups even after taking out the penalties. It can therefore be concluded that explicit training of users on the Web product (i.e. the three colour rules) improves user task performance.

- The experimental design was mainly driven by the explicit condition. Having obtained training on the three colour scheme, users in the explicit condition were expected to make fewer non-optimal responses compared to the implicit group. The experiment has produced a significant difference between the average non-optimal response counts of the two groups (a chi square value of 4.51, $P < 0.05$). It can therefore be concluded that explicit colour learning had a overall positive effect on users' task performance.

- Figure 10.4 and 10.5 show query-wise response times with and without induced delays respectively. As can be seen, in each query response the positive effect of explicit colour training has played a crucial role. This is demonstrated in Figure 10.5 (without delays) where the performance difference between both the groups is large and nearly uniform across all query pages.

- No correlation was found between the response times in Figure 10.5. This indicates that there is no relationship between explicit and implicit test conditions. This could be probably due to the small sample size.

The study has tested the generality issue, i.e. colour design can be applied to A Web page without affecting the original contents, in terms of designing the test task with real pages (see section 10.2.1). It should be noted that the data was not analysed for any evidence of the experimental group learning the rules as it was not part of the experimental design.

In summary, the study demonstrates that people are sensitive to the Web colour product i.e. the three colours rules and associated mappings between the simple perceptual category (colour) and the complex function (type of a link). The study established the effect of explicit learning on users' performance. The effect is robust. The study therefore validates one of the main delivery requirements of the Web product, *mode of learning*: users can explicitly learn the three colour rules.

Chapter 11

Web Experiment 6: Test of Learning

11.1 Foreword

The analysis of results of the previous two studies on the Web have demonstrated that the users are sensitive to colour organisation and they can learn, remember and apply multiple colour rules. In Experiment 4 (Chapter 9), colour rules governing mappings from simple perceptual category, (i.e. colour) to simple cognitive category (selectability of link) were tested. Experiment 5 (Chapter 10) successfully extended this test to complex cognitive categories (selection criteria of links). The results provided strong evidence of explicit learning. Improved users' task performances were consistent with the strategy of training users explicitly on the proposed colour associations. In both the studies, the product as well as the experimental design were driven by the delivery requirement that colour rules/associations would be explicitly described to the users. These studies however were not specifically designed to encourage implicit learning of the colour rules.

The present study, Experiment 6, extends the previous one. The study replicates Experiment 5 in explicit and implicit test conditions. It differs from Experiment 5 in that its design was driven by the implicit learning of the Web colour design. The study intends to investigate whether people could infer the colour rules while performing the task.

The design of the test task was done in two stages. The first part concentrated on formulating a graphic design that allows implicit learning of the colour rules (see Appendix I.2 for design details). The second part implemented the graphic design in designing the Web pages for the test task. The test task was designed to test (a) the generality of the Web product, (b) learning and application of the complex colour associations in both explicit and implicit conditions, and (c) the effect of the colour design on users' task performances.

11.2 Design

The current experiment followed the basic design format of Experiment 5 with some modifications. First, doubling of training time to allow implicit learning of the three colour rules. Second, the test task consisted of nine queries of simple fact retrieval type (see section 9.3). One important aspect of the test task was the use of specially developed nine trial graphic designs to represent hyperlinks in the Web pages (see Appendix I.2).

Hypothesis:

The study was carried out with the hypothesis that: (1) people can learn the three colour rules and associated complex mappings, and apply them across different Web pages; (2) explicit teaching of colour rules is superior to implicit teaching in that it will have a significant effect on users' performance; (3) learning of colour rules, in particular in the explicit mode, can over-ride users' individual biases, such as a left-right bias, location or size bias; and (4) the application of the Web product can be generalised to any Web page.

As part of the test task design, the study used information on some of the Microsoft and Apple products in composing the text and the nine queries. The Web page design included full implementation of the three colour design. As in the previous experiments, the execution of tasks required two types of input from users: (1) text input - in filling forms using a keyboard and (2) mouse input - in selecting hyperlinks.

11.2.1 Design Rationale

The previous two studies have made use of a curriculum vitae on the Web. The present study uses a different set of Web pages in designing the test task. This is intended to examine the generality of applying the Web product to any Web page: a new page to be published on the Web or a page that is already available. The study predicts that using new Web pages in the test task will not affect the application of the Web product.

The Web pages contained details of Microsoft and Apple Products. These products were chosen because of a) the availability of abundant information on these products on the Web; and b) familiarity and popularity of Microsoft and Apple products among users. The aim was to add realism to test task design and create a familiar environment for most users.

The pages used different types of features, for example graphics representing hyperlinks in varying sizes, shapes and location. Both coloured backgrounds and/or borders were applied in designing the graphics. The reason is again to add realism in the design. The Web survey conducted as part of the research (see section 4.4) showed myriad forms of graphics; some were simply images (non-selectable) and the others image maps (selectable). The diversity in graphics design has been replicated in the present study for testing purposes.

Finally, the experimental design and testing was driven by learning aspects of the Web product. In the implicit condition it was possible for users to infer the rules while performing the task. The experimental design consisted of nine trials (see Appendix I.2.1) and was designed to allow learning in the first six trials. The last three trials were intended to test the extent to which the rules had been learned. Also, they included some discordant or antagonistic[1] visual cues i.e. colour vs. shape (see Appendix I.2.1 and I.2.2, Trials 7 and 8), which would provide a strong test of learning and application of the colour rules, in particular by the users in the explicit condition. The reason for adopting such a design was to validate the product from the delivery standpoint: should the rules be taught explicitly or can people infer them?

Operant Conditioning

The rationale for expecting people to infer the rules was based on the concept of *operant conditioning*.

Piaget [Piaget 1970] said about conditioning:

> "Every time you teach a child something you keep him from reinventing it."

During this century, researchers have made attempts to examine this concept through experiments. For example, Thorndike [Thorndike 1913] conducted several experiments with animals as the subjects. He placed a cat in a box and rewarded it with food whenever it escaped. By timing its escape over successive trials, he charted the progress whereby the cat learned to escape to get the food. Initially, the cat was slow in finding the way out by exploration and chance. After some attempts, the cat learned how to escape. Once

[1]A specially designed display feature to over-ride users' learning and application of the proposed colour rules. For example, Trial design 7 in Appendix I.2.1.7. The blue background indicates that the tick mark is not selectable. Despite this visual cue, the tick mark can mislead users into selecting it as the correct link. The presence of tick mark i.e. an antagonistic feature therefore might over-ride the learned colour rules.

learning took place (i.e. the conditioning occurred) it quickly and consistently followed the same plan to find freedom and consequently the food. These experiments demonstrate that the cat's behaviour was conditioned by the environment, by events that took place within that environment and so on [Kaipa 1998]. The studies indicate that even animals can learn to do certain things when they are repeatedly and consistently exposed to the same task.

In the present experiment, the test users were required to select one of the links - blue, green or yellow, repeatedly and consistently across the Web pages. Each colour was associated with a distinct response. Users who selected the yellow link were immediately transferred to the next stage. Green links took more time than the yellow links to do this. Selection of blue links resulted in greater delays followed by error messages. In other words, the users were to be operantly conditioned by the task environment and associated events. Consequently, the implicit group had the opportunity to infer the three colour rules after a few trials.

11.2.2 Method

The test users were randomly assigned to one of the two test conditions. The explicit group served as the 'control group' and the implicit as the 'experimental group'.

All users performed an identical task consisting of nine trials with graphics (see Appendix I.2.1) specially designed for the study. These designs were used in sequence in creating the Web pages. The sequence is important since the graphic trial design was planned to encourage people to infer rules over successive trials. The study predicted that the design would lead to a certain pattern of learning among the users in the implicit group.

11.2.3 Materials

The study was similar to the previous Web experiments in using handouts, the same computer equipment and the Web browser. The following sections describe the visual details of the Web pages used in the test task.

11.2.3.1 Design of the Web Pages

A total of 25 Web pages were designed using the three colour scheme, for both training and testing purposes. Not all the pages were presented to every user; users in the explicit group had to go through more pages than those in the implicit. This was due to an additional colour training session for the explicit group.

The training session included eight pages: five query and three non-query pages. These pages were common to both user groups. They were similar to the pages in the testing session. Also, the training task design involved creating pages for teaching the colour rules to the explicit users. The colour training was similar to that of Experiment 5 (see section 10.2.3.3).

Twelve pages were used in the testing session. Nine were query pages, each containing a distinct graphics design (see Appendix I.2.1). Each page had only one query and 5 to 10 sentences of text in which the answer was embedded. Every non-query page had two or more buttons (Java applets) representing hyperlinks and some text. Target links, their selection criteria (refer to Figure 10.1), and the strategy of adding artificial delays to blue and green links were identical to that of Experiment 5 (see section 10.2.3.1). The golden ratio (see section 10.2.3.1 and Appendix H.1) was employed in the dimensions (i.e. ration of width to length) of rectangular image maps used in query pages 3, 4, 7 and 8 (see Appendix I.2.1 for trial graphic designs). The rest of the test task design was similar to that of the previous study (see section 10.2.3).

11.2.4 Testing Procedure

The testing procedure was identical to that of previous Web experiments. Following the completion of the test task, the users were debriefed and informally asked to express their opinion on the colour design, although subjective evaluation was not included in the testing procedure.

In summary, there were two user groups and each group performed the training and test tasks under one of the two test conditions: implicit or explicit.

11.2.5 Experimental Design

As mentioned in section 11.1, the experiment comprised two sections. The first was paper based and served as the pre-experimental study. The purpose of this stage was to develop trial graphic designs for the second stage. The design details, results and conclusions of this stage are described in Appendix I.

In the second stage, the graphic designs were used in the test task, which was carried out on the computer. Each stage used different test users. Data collected through automatic logging procedures (see section 9.3.3) was analysed in order to obtain the learning and usage of colour rules by each user.

11.2.6 Test Users

Sixteen Liverpool University computer users aged between 19 and 35 served as test users. Seven were female. None had participated in previous experiments testing colour. All had normal or corrected to normal vision. The Web experience of users varied from eight months to one year and they were randomly assigned to the two test conditions, with eight users in each.

11.3 Results

The data was subjected to statistical analysis. The main performance measures were task completion times and non-optimal response counts (see Figures 11.1 and 11.2) as described below.

11.3.1 Performance of Tasks: Completion Time

The primary measure was the time taken to complete the test tasks in each condition. The mean completion times for both explicit and implicit groups were 461.1 and 796.4 seconds, respectively. These times included artificial delays as described in section 10.2.3. The mean completion times without delays for the groups were 381.8 and 522.8 seconds respectively. The mean completion time for each test group is shown in Figure 11.1.

A Mann-Whitney U test (see section 6.2.7) is used to determine whether there are any significant differences between the time scores of the two test groups. The test is one-tailed. The test results indicate a significant difference between the time scores of both the groups with (Mann-Whitney U value of 0; probability less than 0.001) and without delays (Mann-Whitney U value of 12; probability less than 0.05).

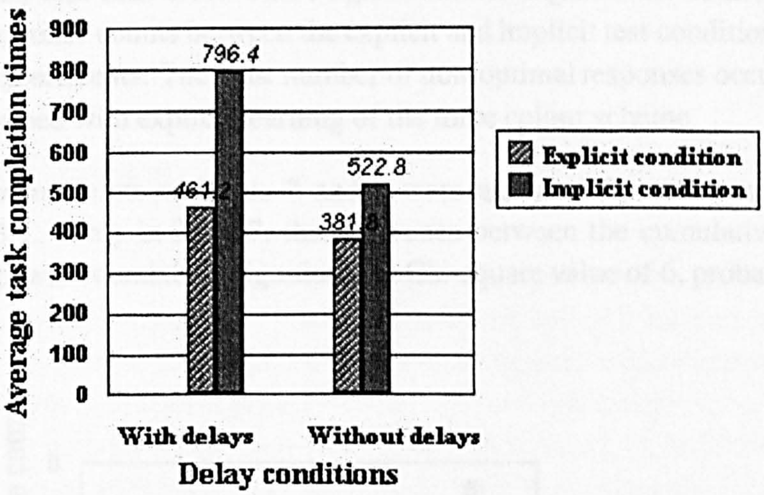


Figure 11.1: Test conditions and average task completion times in Experiment 6.

The trial-wise response times, with and without induced delays, are shown in Figures 11.3 and 11.4. In particular, the completion times of Trials 7 and 8 (after removing delays) were analysed since they include antagonistic features, which were expected to provide a strong test of explicit learning. A significant difference was found between the response times (see Figure 11.4) of Trial 7 (Mann-Whitney U value = 13, P = 0.025) indicating the implicit group performed better than the explicit. In the case of Trial 8, no significant difference was observed between the completion times (Mann-Whitney U value of 19; P = 0.097). This indicates a similar level of performance in both groups.

Further, Spearman's rank correlation coefficient is calculated for the explicit and implicit query response times in Figure 11.4. The test indicated that there is no correlation (correlation coefficient = 0.43, N = 9) between the response times.

11.3.2 Non-optimal Responses

The average non-optimal response counts for the explicit and implicit groups were 1 and 8 respectively (see Figure 11.2). A Chi-square test was used to analyse non-optimal response counts. The results show a significant difference with a chi-square value of 5.44 and a probability less than 0.05. These figures further augment the consistent pattern of non-optimal response counts between the explicit and implicit test conditions found in the previous two experiments. The least number of non-optimal responses occurred when the task was performed with explicit learning of the three colour scheme.

The non-response counts of Trials 7 and 8 were analysed for the reasons mentioned in section 11.3.1. Only in Trial 7, the difference between the cumulative non-optimal response counts was found to be significant (a Chi-square value of 6, probability less than 0.05).

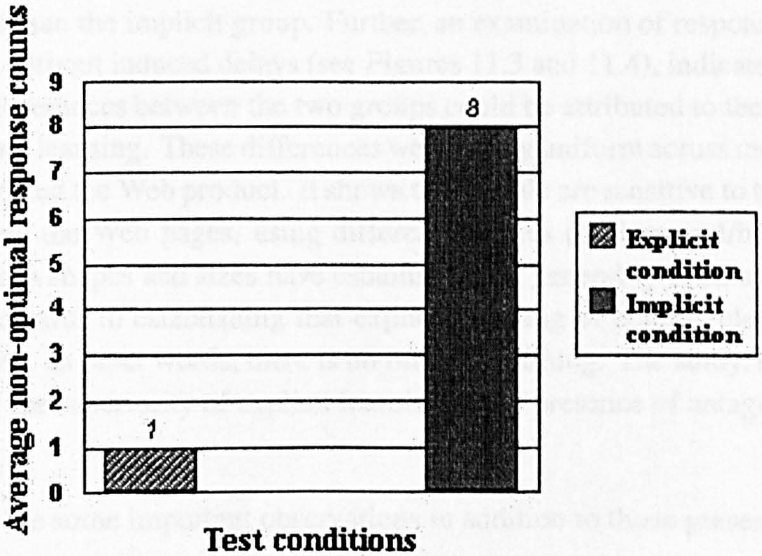


Figure 11.2: Test conditions and average non-optimal response counts in Experiment 6.

11.3.3 Debriefing

As in the previous studies, all users were debriefed and were invited to comment on the three colour scheme and the study itself after they had carried out the test. Users' opinions about the small number of rules and the novel use of colour on the Web were almost identical to that of Experiment 5. In this study, some of the graphic designs received special attention from some users. Seven users described Trial 7 as misleading because the design (see Appendix I.2.1.7) encouraged the users to select a non-target region of the image map (i.e. the tick mark). Three users described Trial 8 as rather interesting because of its shape (see Appendix I.2.1.8) The test users were briefed about the rationale for using different graphic designs.

11.4 Conclusions

Following the hypothesis (see section 11.2), the explicit group was expected to produce a better performance than the implicit group. Figures 11.1 and 11.2 indicate that the explicit group not only took less time to complete the task, but also recorded a smaller non-optimal response count than the implicit group. Further, an examination of response times across trials with and without induced delays (see Figures 11.3 and 11.4), indicated that the high performance differences between the two groups could be attributed to the positive effect of explicit colour learning. These differences were nearly uniform across most of the trials. The study validated the Web product. It shows that people are sensitive to the three colour rules. Changing the Web pages, using different features (background/border), colours, images in various shapes and sizes have established the generality of its application. The study was successful in establishing that explicit learning of colour rules is superior to implicit learning. In other words, there is no implicit learning. The study, however, failed in establishing the superiority of explicit learning in the presence of antagonistic features (see below):

The following are some important observations in addition to those presented in sections 9.5 and 10.4:

- In Trial 6, task response times of both the groups were nearly the same (see Figure 11.4). The users in the explicit group took more time than expected. This trial used border colour as the target feature. It is probable that the border feature was poorly learnt by the explicit group which in turn could be due inadequate training on this colour feature.
- It should be recalled that, Trials 7 and 8 were specially designed to have antagonistic features (see Appendix I.2.1.7 and I.2.1.8). The purpose was to examine whether

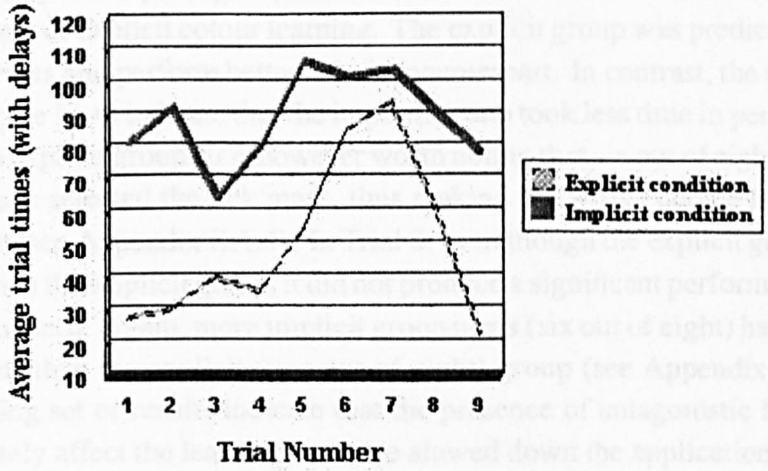


Figure 11.3: Test conditions and average trial completion times (including delays) in Experiment 6.

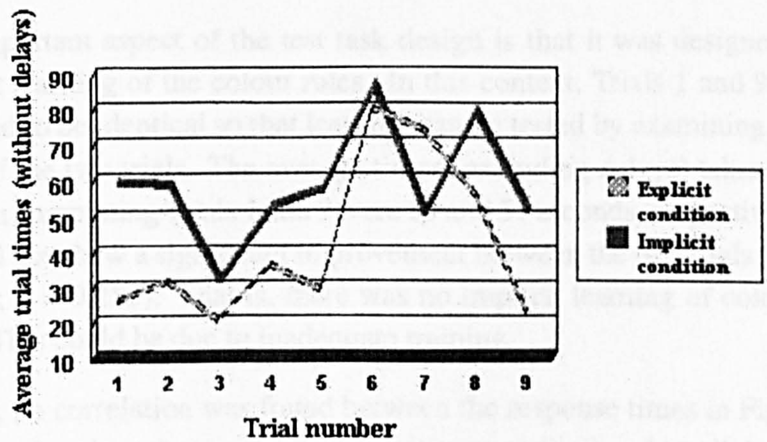


Figure 11.4: Test conditions and average trial completion times (excluding delays) in Experiment 6.

---

these features could over-ride the learned colour rules, in particular, the learning of the rules by the explicit group. In other words, the two trials were assumed to be a strong test of explicit colour learning. The explicit group was predicted to apply the colour rules and perform better than its counterpart. In contrast, the analysis of data (see Figure 11.4) indicate that the implicit group took less time in performing Trial 7 than the explicit group. It is however worth noting that six out of eight implicit group users have selected the tick mark, thus making the wrong choice first followed by the right (see Appendix K.1.4). In Trial 8, even though the explicit group performed better than the implicit group, it did not produce a significant performance difference between them. Again, more implicit group users (six out of eight) have made wrong decisions than the explicit (two out of eight) group (see Appendix K.1.4). These interesting set of results indicate that the presence of antagonistic features did not completely affect the learning but have slowed down the application of rules by the explicit group. It can therefore be concluded that some features on the display, such as the antagonistic features used in the trials, can affect or over-ride the application of the learned colour rules. This could be due to either poor learning of colour rules by the users or a conflict between learned rules: that is "tick" is good is a learned feature. Also the users' model of the Web colour product may be cognitively weak relative to other models that they have about learning and applying rules. Although the implicit group has yielded better task completion times, the presence of errors indicate that there is no evidence of good learning of rules.

- One important aspect of the test task design is that it was designed to encourage implicit learning of the colour rules. In this context, Trials 1 and 9 were specially designed to be identical so that learning can be tested by examining the completion times of the two trials. The average times (excluding delays) taken by the implicit group in completing Trials 1 and 9 were 59 and 51 seconds, respectively. Analysis of data did not show a significant improvement between the two trials (Mann-Whitney $U = 25$; $P = 0.253$). That is, there was no implicit learning of colour rules by the users. This could be due to inadequate training.

- Further, no correlation was found between the response times in Figure 11.4. This indicates that there is no relationship between explicit and implicit learning conditions; this could be due to small sample size for obtaining a significant correlation.

In summary, the study demonstrates that people are sensitive to the three colour rules and associated complex mappings. The failure of the study in establishing implicit learning of the rules indicates that the proposed three colour scheme (and similar schemes) need to be taught explicitly to users. Explicit teaching is therefore considered as an important delivery requirement of the Web colour product.

Chapter 12

Web Experiment 7: Redesign & User Evaluation

12.1 Foreword

Previous studies (Chapters 9 to 11) used different experimental designs to validate the Web product. In addition to performance measures, for example task completion times and non-optimal response counts, generality of the product and its acceptance by users (i.e. subjective satisfaction) were considered as important delivery requirements. However, a product evaluated favourably on every performance measure, still may not be accepted by users for aesthetic reasons [Chin *et al.* 1988]. The study reported in this chapter was part of the research effort to develop a questionnaire that directs attention to users' subjective rating of the Web colour product. Prior to distributing the questionnaire, a style guide [Sahithi and Diaper 1998] for implementing the product was developed and published on the Web for the purpose of user evaluation (see Appendix J). The style guide involved applying the product to some Web pages already available on the Web. In addition, the style guide demonstrated the implementation of the product on some real Web pages, for example Web pages of Bournemouth university, UK. The following sections discuss the style guide and the study in detail.

12.2 Style Guide for the Web Product - The Three Colour Scheme

The style guide encompasses (a) the broad principles, which offer general advice, (b) provides a rationale for the proposed design, and (c) the specific design rules concerning the Web colour scheme and associated novel colour functionality (see section 4.5). The style guide was intended primarily for a diverse population of users, who are familiar with Web navigation concepts and graphic browsers, such as Netscape or Internet explorer. It seeks to achieve consistency in functionality [Benyon *et al.* 1990] both within and across Web pages through the application of the colour scheme.

The style guide was organised into seven sections as shown in Figure 12.1. The figure shows seven nodes fully connected via hyper-links. (connectivity not completely shown for the sake of clarity). Each node represents a different section in the style guide. The home page was primarily an introduction to the contents and general structure, and includes a request for feedback on the proposed colour scheme and the style guide itself. Branching out from the home page are the other hyper-linked sections:

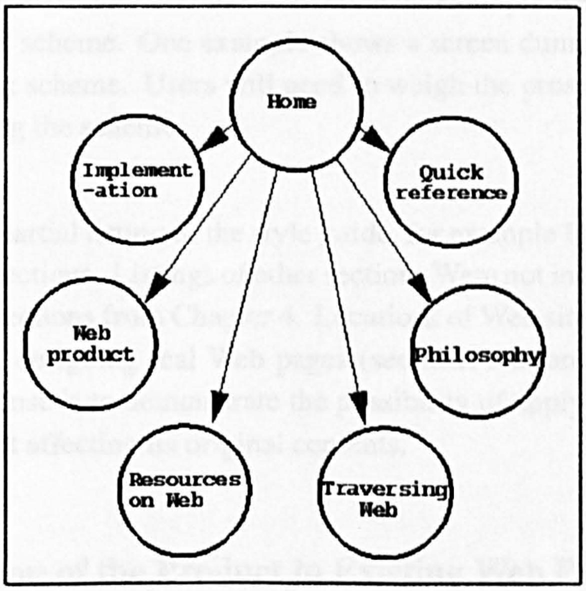


Figure 12.1: Style guide for the Web product. All pages are fully connected through hyperlinks. The figure shows connectivity with the Home page only.

- **Quick reference**: This section highlights the rudimentary aspects of the Web colour scheme as described in Chapter 4: classes of resources, representation of classes using colour, target features (background/ border colour), rules and associated colour functionality.

- **Philosophy**: The section includes background information, on the purpose of the proposed Web colour product.

- **Traversing the Web**: The purpose of this section was to introduce the role of hyperlinks in a Web page and different ways of representing them. The goal was to inform users about a variety of forms of representing hyperlinks (e.g. text, image maps, buttons etc.).
- **Resources on the Web**: This section introduces types of Web resources (e.g. text, graphics, video clips etc.) and associated problems in selecting them. The section reproduces the proposed classification of resources discussed earlier in Chapter 4.
- **Three colour scheme**: This section is a short version of section 4.5 and provides details of the web colour scheme.
- **Implementing the colour scheme**: This contains example Web pages implementing the three colour scheme. One example shows a screen dump of a page redesigned using the colour scheme. Users will need to weigh the pros and cons of designing Web pages using the scheme.

Appendix J includes partial listing of the style guide, for example Home page, Philosophy and Implementation sections. Listings of other sections Were not included as they replicate details of associated sections from Chapter 4. Locations of Web sites, which implemented the colour scheme in designing real Web pages (see next section) were provided in the style guide. The purpose is to demonstrate the possibility of applying the colour scheme to a Web page without affecting its original contents.

12.2.1 Application of the Product to Existing Web Pages

A few pages present on the Web were redesigned using the three colour scheme. The goals of the redesign efforts were to [Nielsen 1998]:

- Facilitate easy identification of selectable and non-selectable areas of the Web display, *at a glance*.
- Provide a unified visual appearance of heterogeneous resource content, *at a glance*, without actually traversing hyperlinks.
- Reduce search and/or browsing time and consequently cognitive load and negative emotions associated with traversing unimportant/less desirable links.

- Ensure high usability and quality user experience.

Appendix J.1.4 consists of screen dumps of some real Web pages, before and after applying the web colour scheme. The latter illustrates some immediate advantages of the Web colour scheme. For example, it allows users trained on the colour rules to visually distinguish selectable image maps from non-selectable ones, *at a glance*.

12.2.2 Implementation of the Product in Real Web Pages

During the course of the research reported in this thesis, the author had an opportunity to redesign some of the Web pages of Bournemouth University using the three colour scheme. In assigning the colour codes, both accessibility and desirability have been considered following the scheme proposed in Chapter 4[1]. Redesigned pages were published at Bournemouth university site (http://dec.bournemouth.ac.uk/) for a period of eight months (not available on Web since 10th November, 1998). Screen dumps of sample Web pages are provided in Appendix J.2. People who visited the site have sent useful feedback (discussed in next section) on the application of the colour design.

12.3 Questionnaire

> "It's not just the utility of the program that matters, it's also the user's acceptance of the system – subjective satisfaction - that is a crucial measure of a program's success" - [Norman 1990b].

Shneiderman [Shneiderman 1998] says, " User satisfaction is considered a diagnostic of the strengths and measure of strengths and weaknesses". This applies to the proposed Web colour scheme. Following the literature review [Slaughter *et al.* 1995, Chin *et al.* 1988, Harper and Norman 1993, Vanniamparampil *et al.* 1995], an empirical study was carried out to investigate the use and perceived impact of the Web scheme using a questionnaire. The questionnaire was developed in order to gauge the overall subjective reaction of a

[1] All links were non-local (score 2) with unrestricted access (score 3). Load times were based on file sizes. This has resulted accessibility values of 6,7 or 8. In assigning desirability value, some of the factors considered were text-only, text and graphics, graphics-only, HTML pages with forms and menus. The final score for the links ranged from 7 to 11. The colour codes were assigned based on this score.

user to: (i) the proposed novel functionality of colour in Web page design; and (ii) the implementation of the product in designing/redesigning real Web pages. The goal was to assess user satisfaction with respect to understandability, learnability and applicability of colour, as an important and useful visual cue.

The questionnaire consisted of twelve questions as shown in Table 12.1. Nine were of judgmental type (questions 1, 2, 4-6 and 9-12). Before they give a satisfactory rating for these questions, users were required to analyse the three colour scheme and its implementation aspects sufficiently to formulate an opinion based on it. Three questions (3, 7 and 8) were overall reaction ratings to the colour scheme. The table includes some of the questions listed earlier in Table 7.4. The questionnaire was based on a format reported by Spool *et al.* [Spool 1997]. Both paper and online versions of the questionnaire were prepared for distribution among test users.

The responses were rated on a 7 point scale: positive adjectives anchor the right side and negative adjectives anchor the left. Each question was designed to measure the users' overall satisfaction with respect to different aspects of the colour scheme.

12.3.1 User Groups, Materials and Methodology

The study consisted of two user groups. In the first, the questionnaire was offered to users who have participated in previous experiments testing the Web product. The group consisted of twenty test users, twelve males and eight females in the age group 20-26 years. They differed widely in their level of computer or Internet experience. The evaluation took place with one user at a time. Each user revisited some of the experimental Web pages (online or screen dumps) and attempted to remember the three colour scheme and associated functionality. Test users were given a simple demonstration of the colour scheme along with instructions describing how to complete the questionnaire. Following the instructions, the questionnaire was distributed along with data sheets, one per user, and their ratings were recorded.

In addition to paper-based questionnaires and data sheets, electronic copies of the questionnaire were distributed to twenty Web users outside the UK. Each of these users, aged between 25-40, were contacted individually. None had previous experience of colour studies or knowledge of the proposed novel functional allocation of colour on the Web. Only six of them, five males and one female responded to the author's request to fill-in the questionnaire, thereby forming the second user group. Along with the details of the three

1	How do you feel working with consistently coloured hyperlinks? *completely confused* ——— **1 2 3 4 5 6 7** ——— *everything made sense*
2	How sensible did the functional allocation of colour seem to you? *unacceptable* ——— **1 2 3 4 5 6 7** ——— *Excellent*
3	Rate the ease of identifying different types of resources *very difficult* ——— **1 2 3 4 5 6 7** ——— *very easy*
4	How easy was the colour design to learn and remember? *very difficult* ——— **1 2 3 4 5 6 7** ——— *very easy*
5	How easy was it to adopt them to displays that implement colour design? *very difficult* ——— **1 2 3 4 5 6 7** ——— *very easy*
6	How appropriately were the colours used? *very unsatisfied* ——— **1 2 3 4 5 6 7** ——— *very satisfied*
7	Rate overall ease of use *very difficult* ——— **1 2 3 4 5 6 7** ——— *very easy*
8	Rate your overall productivity with the scheme *very unsatisfied* ——— **1 2 3 4 5 6 7** ——— *very satisfied*
9	Would you recommend designing pages using the three colour scheme? *Not at all* ——— **1 2 3 4 5 6 7** ——— *very much*
10	Does the three colour scheme affect your aesthetics? *Not at all* ——— **1 2 3 4 5 6 7** ——— *very much*
11	Did you appreciate redesigning Web pages using the proposed scheme? *Not at all* ——— **1 2 3 4 5 6 7** ——— *very much*
12	While performing test tasks that implemented the colour scheme, did you: *feel completely frustrated* ——— **1 2 3 4 5 6 7** ——— *always know what link to select next*

Table 12.1: Questionnaire for studying users' attitude towards the three colour scheme. Responses were counted on a 7-point scale.

colour scheme, instructions describing how to complete the questionnaire were distributed through email and each user was requested to browse through the style guide (see section J.1) before attempting to fill-in the data sheet. This electronic form did not include the last question listed in Table 12.1 because it was only relevant to the first group of users who took part in the previous experiments.

In addition to the above two groups, four more highly qualified computer science professionals, one from the University of Liverpool and three from the University of Bournemouth, were invited to participate in the study. They did not fill-in the questionnaire, but have sent their descriptive opinions about the Web colour scheme. Consequently, these users were treated as a separate group and their opinions were described in the next section.

12.3.2 Users' Feedback: Examples

The first approach to the analysis of the results is to calculate the mean rating for each question for each group. The profile of mean ratings and the overall mean rating for user groups 1 and 2 are shown in Figures 12.2 and 12.3, respectively.

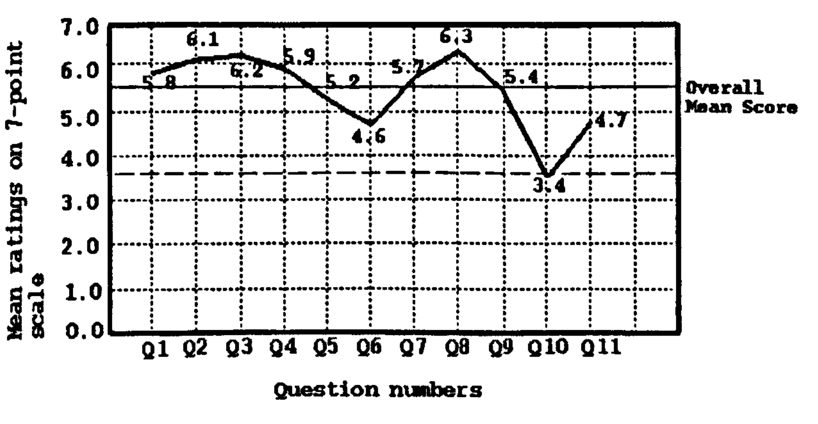


Figure 12.2: User group 1 - Users' evaluation of the three colour scheme - Mean ratings for each question on a 7 point scale in Experiment 7. The solid black line indicates the position of the overall mean rating (5.4). The dotted line shows the position of neutral rating (3.5).

The second approach to the analysis of ratings is to determine the three highest and three

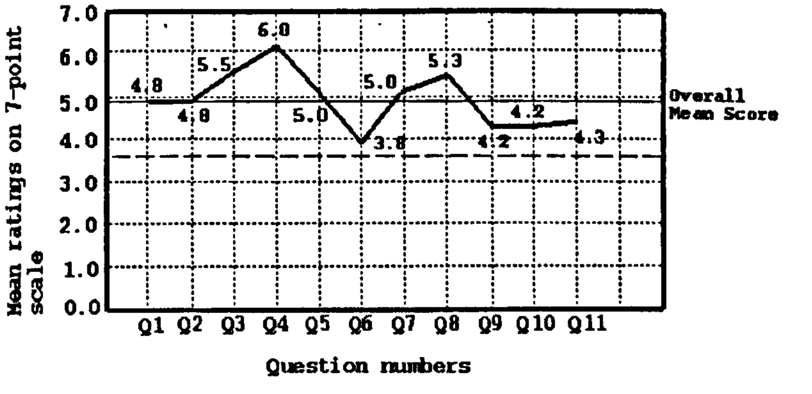


Figure 12.3: User group 2 - The figure shows how users rated the individual questions referring to various aspect of the Web colour scheme on a 7 point point scale in Experiment 7. The solid black line indicates the position of the overall mean rating (4.8). The dotted line shows the position of neutral rating (3.5).

---

lowest rated aspects of the colour scheme. The results are summarised in Table 12.2.

The third approach to the analysis is to summarise comments from users, in particular, respondents who did not complete the questionnaire but have mailed descriptive opinions. A brief discussion follows each comment.

Description of Users' Comments:

"The choice of colour seemed arbitrary but I very soon got used to which meant what. On the pages accessed, the additional colour caused no confusion although I would imagine that on inherently more colourful pages it may do."

"I am being as critical as possible. I think the basic idea is excellent, a useful piece of additional information."

Comments made about the proposed novel functionality and use of colour as the vital visual cue were nearly the same. A medical student in the first group concluded *"Since there are only three rules, I really did not have that much trouble learning and remembering the colour scheme. More over, I like the way colour has been used, instead of puzzling icons, to convey helpful information."* Insufficient information was reported by two users in the

User group	Question (Rating) Highest three	Question (Rating) Lowest three
First	Q8 (6.3) Q3 (6.2) Q2 (6.1) *	Q11 (4.7) Q6 (4.6) Q10 (3.4)
Second	Q4 (6) * Q3 (5.5) Q8 (5.3)	Q11 (4.3) Q9 *, Q10 (4.2) Q6 (3.8)

Table 12.2: Questionnaire - Highest and lowest rated questions common to both user groups addressing different aspects of the Web colour scheme on a 7 point scale in Experiment 7. The star symbol indicates questions which were not put into the same category by both the groups.

---

second group. These users asked for more examples.

"An interesting notion. I tried your path through the pages, and points from my subjective impression are as follows

- *I kept forgetting that the colours made any difference, though with time it/I got better.*
- *I felt that green would be better as the 'most interesting' colour. I find Yellow rather intrusive and I found myself tending to scan the green ones for 'good stuff', and having to make myself scan the yellow ones.*
- *It certainly gave a useful structure to the use of colour, and I think I would like it.*
- *Not trying to be flippant, but I am not sure that 3 colours is not too many in one sense: I found myself ignoring the blue. I am rather impatient with web pages, so this may be a very individual remark."*

Statements included suggestions for interchanging the role of the colours, and a new colour set such as red, green and blue. While making a suggestion against using yellow, users made a point of stating that they often can not see some shades of yellow against lighter page backgrounds. Suggestions made about the shades of colours also varied. Users' preference for different colour sets was reflected in the ratings of questions 6 and 10.

"I'm afraid I don't find the colours at all helpful. The distinction between green and yellow seems to be entirely arbitrary (as a user). I guess it mainly depends upon the length of the html file. The problem is 4 pages of html isn't much of a problem. Perhaps you might flag large graphics files but even then there are still 2 unknowns (i) network speed (ii) server load. Blue seems to indicate the link can't be reached in which case why put it in the web page in the first place? (unless of course it is dynamically updated (every few seconds?))

Although the graphics or design is pleasing to the eye, I'm not terribly impressed with the use of colour. May be you should use rather more shades? It's a bit coarse at the moment."

Comments made about dynamically updating the status of linked-to resource content suggest the need for developing a plug-in like system. The main features of such software are outlined in section 13.3.5.

12.3.3 Results

Overall rating: The overall mean rating of the questions (first eleven listed in Table 12.1) for the first and second user groups are 5.4 and 4.8, respectively. From the profile of results shown in Figures 12.2 and 12.3, it is obvious that the proposed web colour scheme was rated well on a majority of aspects. If five, on a seven-point scale is to be defined as an acceptable overall rating of a "good" design, then the overall rating made by twenty users in the first group shows that the three colour design is successful in meeting the delivery requirements - simplicity and consistency. Overall rating in the second group is close to target rating i.e. 5 and therefore is considered as a measure of acceptance of the Web product. The difference between overall ratings, however, indicates that the users who participated in previous colour studies showed better appreciation of the colour design relative to those who absorbed the colour details by looking at the style guide. The experience gained in performing representative test tasks therefore had a beneficial effect on users rating. In addition, users in the first group had given the last question (see Table 12.1) a mean score of 6.8, the highest of all ratings. This rating clearly indicates that users had a good understanding of the product while performing test tasks and after a quick reminder, they were comfortable in answering the questionnaire after a period of four to seven months.

Highest and lowest rated questions:

An important examination of the results was an analysis of the three highest and lowest rated questions.

Questions 3 and 8 (common to both the groups), and 2 and 4 had a high rating because they directly relate to users' performance in using the scheme. All the questions with highest ratings had a mean rating of more than five and represent the best aspects of the scheme.

Questions 6, 10, 11 (common to both the groups) and 9 were rated low, each with a mean rating less than five. Among the three, question 10 received the lowest rating. It was concerned with the aesthetic aspects of using the three colours. Users seem to have different preferences for colours other than those in the proposed set. As mentioned in Chapter 4, the research suggests that the use of other possible sets of colours need to be explored in implementing the novel functionality (see scope for further research in Chapter 13. The same explanation is applicable to Question 6, which received the second lowest rating. Despite the effort to employ browser safe colours [Weinman and Heavin 1997b], a majority of users might have failed to see the actual shades of colours due to incompatible hardware. Forcing colour control, that is, over-riding users' set of colour preferences (such as using a different colour map), is quite difficult on different window platforms. In particular, some shades of yellow tested in the study on a HP-Unix platform appeared in pale-white on IBM PCs. This could be one of the main reasons for receiving low ratings with respect to design-redesign issues using the proposed colour set. These aspects could further be improved to cater to the needs of users, including aesthetic aspects.

Overall, users in each group shared nearly the same opinions regarding the Web colour scheme and its role in designing Web pages. Consistent with findings in previous experiments (Chapters 9 to 11), the overall ratings clearly indicate that the colour design had a beneficial impact on users' assessment and attitude towards learning and applying it. In particular, users in the first group, who had participated in the empirical studies and acquired hands-on experience in using pages that applied the product, were found to be more confident and well organised in answering questions. Furthermore, these users found the consistently coloured Web pages informative and helpful. Comments made by the users during the questionnaire session also reflected similar opinions. In summary, users were found to be sensitive to the colour organisation using the Web colour scheme.

Using the questionnaire in the evaluation of the Web colour scheme helped in assessing its strengths and weaknesses. The comments collected from the questionnaire provided a list of suggestions, complaints and endorsement for the colour product. These results are helpful for further research and provide a realistic evaluation of the proposed use of colour.

Chapter 13

Conclusions and Scope for Further Research

13.1 Foreword

In the preceding chapters, a colour product for functional Web page design has been presented. The product is a three colour scheme for colour-coding the hyper links based on resource types. The scheme has been developed within the research framework described in Chapter 1 (Figures 1.1 and 1.2). The design of both the product and the experiments was constrained by delivery requirements.

The Web product was designed using colour as an important physical cue. Colour is one of several functionally under utilised features in GUI designs and the research focused on using colour for enhancing the functionality of Web page designs. A series of experiments were conducted on the Web (Chapters 9-12) to evaluate the product with respect to the main delivery requirements: learnability, simplicity and consistency.

The first section of this chapter summarises the main contributions of the work reported in the thesis. The second section proposes topics for future work.

13.1.1 Research in Retrospect

In recent years high quality colour displays became more affordable making their usage ubiquitous. Colour is a powerful visual cue, already available, now cheap. When used properly, colour is considered generally good and superior to other visual cues, such as size and shape, for conveying potentially useful information to users' during task performance (Chapter 2 for more details). Unfortunately, colour is one of the poorly understood subjects amongst interface designers. The research presented in this thesis argues that colour is an under-exploited visual feature (Chapter 1) relative to features such as menu design and icon appearance [Diaper and Sahithi 1995]. The extent of research done on exploiting the functionality of colour is by and large limited to crude grouping and the signalling of emergency states. Further research on this feature was desirable and suitable for a doctoral programme.

Consequently, the research has focused on investigating novel potential uses of colour hitherto unexplored. The goal was to explore what additional information, that was not already generally available on most computer systems, can be communicated by using colour. The Web was considered as a suitable platform for the purposes of investigation and delivery.

This research involved developing two products i.e. REPs (see section 1.5.2) within the thesis framework (see Figures 1.1 and 1.2). First, a product was designed for traditional UIs to investigate the use of colour (see Chapter 3). This was intended to serve as a pilot product. The associated experiments (see Chapters 6,7 and 8) formed the test bed for the Web colour product. Design of both the product and pilot experiments was driven by the delivery requirements. The first product design involved using coloured background or border features consistently to convey information about screen areas that accept different types of input, such as keyboard and mouse, *at a glance*. For example, the design was intended to allow users to rapidly identify selectable screen widgets and/or areas from non-selectable ones. Initially, more than a dozen application interfaces were examined on a Unix platform for implementing the colour design. Many of them were used to try and demonstrate the advantages of the colour design but not all of them were found suitable because of the technical difficulties in changing individual widget colours. Prior to experimental evaluation, applications such as Xscribble and Xbiff, were colour coded using the proposed colour set for user feedback. Later, the colour scheme was subjected to empirical evaluation (see Chapters 6, 7 and 8) in order to investigate its effect on users' task performance under varying test conditions. Each experiment used different applications in order to test the generality of applying the product. Experiment 2 provides, by its failure, a caution and a justification. The caution is that an apparently simple and straightforward design for colour user interfaces could not always be applied usefully to existing software. It justifies the careful empirical approach to the design and delivery of the Web colour scheme.

The advent of Web with interconnected resources and selectable hyperlinks has brought a new level of sophistication to the Internet. Unfortunately, not all information on the Web is easily accessible. Where the graphics content is high, other factors such as downloading time play a critical role. Despite its advantages, navigating millions of pages currently available on the Web present myriad of problems in accessing desirable information [Dillon 1991, Edwards and Hardman 1989]. For example, selecting a resource with restricted read permissions (e.g. http:// www.liv.ac.uk/ Library/ news) or a statistical dump of numbers (e.g. http:// syy.oulu.fi/ log.html), are commonly reported by users (Chapter 4). Often there are no visual cues to let a user : (1) visually distinguish selectable regions from non-selectable ones within a page without moving the cursor; and (2) assess the contents of linked-to resources without actually traversing the link.

The thesis has designed and delivered a product addressing the above problem. The high level colour design is a major part of the thesis. The creative part of the design involved using background or border features of colour consistently to indicate resources (i.e. information) with different degrees of accessibility and desirability. Initially, a Web survey was conducted to identify different types of resources available on the Web. Following the survey, a coding scheme was proposed (see section 4.4) to make an assessment of a given resource based on its accessibility and desirability. The assessment yielded a value to each resource. The resources were classified based on their values and colours were used to represent the classes. The number of colours (consequently, categories) was constrained by the need for simplicity of the product, a delivery requirement. The product design was restricted to only three colours, blue, green and yellow, so as to make it easy for users to learn, remember and apply the associations between colour and the type of the resource or link. The coding scheme yields a value, which is a combination of overall accessibility and desirability. Yellow indicates resources with high degree of accessibility and desirability. Green signals an intermediate level and blue represents resources with low or zero accessibility and desirability. Red is reserved for signalling emergency states or problems of a serious kind. The three colour product is explained in detail in sections 4.4 and 4.5 along with suitable examples.

Application of the Web product allows the users who know the design conventions to immediately identify the types of links in a Web page, *at a glance*, i.e. without traversing the link or moving the cursor. For example, blue colour links, immediately indicate that they point to the least desirable resources, such as data dumps, dead pages or under construction pages. Such information is very helpful, especially in time critical situations where a user has to browse through large numbers of pages.

The three colour product was subjected to empirical evaluation and subjective assessment (Chapters 9 to 12). The experiments were constrained by the delivery requirements on learning(explicit vs. implicit). Further, the experiments used different Web pages in order to test the generality of applying the product as it was on this issue that the UI colour design failed (i.e. Experiment 2). The results with regard to the novel functional allocation of colour were consistently favourable indicating that the product had beneficial impact on users' performance as well as on their assessment. The experiments established that people are sensitive to colour organisation and explicit learning of the colour rules produces better user performance. The application of the colour scheme across different Web pages produced similar results, thereby establishing its generality. On the other hand, users' were not completely satisfied with the selection of colours (and shades) employed (see sections 12.3.2 and 12.3.3). Evaluating different colour sets and associated psychological and physical aspects is beyond the scope of this thesis. Suggestions from users about using different colour sets and shades were presented for the purpose of future research.

In summary, the research presented in this thesis could be described successful in achieving the overall objective of the thesis (see section 1.1). The aims of the research have been met in terms of proposing a plausible design solution for the Web (Chapter 4) and evaluating its usage (colour bindings, learning and remembering colour rules) by conducting a series of experiments (Chapters 9 to 12). It has been established that, as a vital visual cue, colour can be used in a novel way to signal information, *at a glance*, in a realistic environment (i.e. on the Web platform). The important usability factors that must be taken into account when designing, implementing, and delivering the three colour scheme were explored. Empirical evaluation of the Web colour design demonstrates that it is feasible to implement it consistently, both within and across pages.

13.2 Contributions Of Work

HCI research: The basic premise of this thesis is that HCI research should be about methods that help users (both designers and end users) to create and maintain functional, consistent and aesthetic work environments. The research identifies important display features that appear to have been under exploited and demonstrates the need for further research by taking colour as an example. Furthermore, it draws attention to the situation that HCI knowledge, and WPD and UID practice are inadequate at present with respect to the use of colour in user interfaces. While exploring different aspects of colour, it emphasises the need for developing functional and consistent visual displays to assist users during task performance.

Colour has potential uses: The thesis demonstrates that it is possible to assign novel functionality to colour such that users can easily associate a specific function with a specific colour. Furthermore, it confirms that users can easily learn, remember and apply colour rules.

Rapid extraction of information is important: The research illustrates the possible benefits, such as improved user performance, from designing displays that allow users to extract information rapidly (e.g. *at a glance*) in performing tasks, for example Web browsing. The work was less successful on traditional stand-alone user interfaces.

Delivering products: Delivering products, simple, high-level plausible designs, to Web users who are not experts in HCI and usability issues, is an important contribution of this thesis. Analysis of data and users' ratings confirm the validity of the proposed Web product and thereby provides empirical support for designers who wish to consider applying it to Web pages.

13.3 Topics for Future Work

The following sections briefly describe possible topics for future research.

13.3.1 Testing More Real Web Pages

Designers are most likely to use a design if its validity is tested extensively from a wide variety of aspects. Despite the evaluation of the three colour scheme in the thesis, the research could not test the colour scheme on real Web pages in sufficiently large numbers. This may be required to provide a stronger evidence base to the Web designer community. Usability testing could be extended further to and evaluate the Web product in designing or redesigning real Web sites.

13.3.2 User Feedback

Users criticisms about the colour set and shades of colours are not currently handled in this thesis. Colours that a user has expressed to be more appropriate than the current colours need to be analysed and used to increase user acceptance of the Web scheme. The empirical approach for doing this, however, needs to be carefully designed, so that

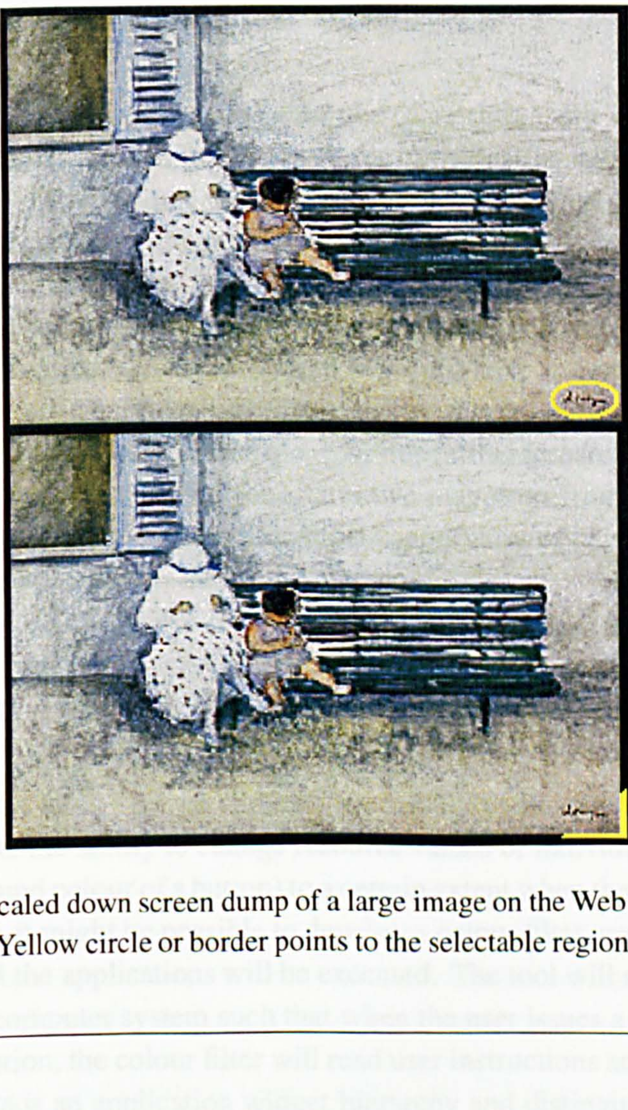


Figure 13.1: A Scaled down screen dump of a large image on the Web with colour design implementation. Yellow circle or border points to the selectable region in this image map.

suggestions regarding new colour combinations meet users recommendations. A feasible approach is to create some displays based on users' choices and testing them against other combinations of colours [Macintyre 1991, Chen *et al.* 1990].

13.3.3 More Colour Functions

Another area of improvement could be to use more than three colours proposed in the Web scheme. Each class in the present resource classification combines several attributes of the resource in order to limit the number of colours to three. The number of classes could be increased further, perhaps to six or even nine, and different shades of each colour or a whole new set of distinct basic colours could be used to represent classes. The cognitive overhead on users to learn and remember to associate more functions to colours needs careful evaluation. In the proposed Web product, a simple perceptual category colour is mapped onto a complex function of different types of resources. The complexity associated with resource classification also requires further research. The proposed scheme is simplistic and may be explored for alternative mappings from functions onto colour. Future research may involve finding the most useful ways of characterising accessibility and desirability to provide a basis for the mapping.

13.3.4 Colour Filter: A Tool for Traditional UIs

Currently available tools such as "Editres" (a resource editor on Unix platform; section 2.5.2) demonstrate the ability to change resource values of individual application widgets (e.g. the background colour of a button) to a certain extent when the application is running. Given this ability, it might be possible to develop a colour filter, probably a super-Editres, through which all the applications will be executed. The tool will sit somewhere between the user and the computer system such that when the user issues a command to execute a particular application, the colour filter will read user instructions and invoke the requested application. It reads an application widget hierarchy and distinguishes widgets that take inputs (keyboard, mouse etc.), from those which are simply labels (e.g. non-selectable widgets such as Xbiff in Figure 3.1). The tool then reads background, foreground and border values of each widget and modifies them to comply with the proposed colour design. The tool will change the foreground colour for a proper contrast and runs the application with new widget colour values. Further investigation could be made towards developing the systems that support such tools [1].

13.3.5 A Plug-in for the Web Product

An advanced version of a colour filter is required to implement the Web colour scheme. This version could be a plug-in that runs inside a graphic Web browser, such as Netscape Navigator or Internet Explorer (ActiveX component). Before displaying the Web page to the user, the tool will analyse the Web page contents and interlinked resource types by employing caching and prediction procedures [Saunders *et al.* 1998]. Once the link regions and linked-to resource types are identified, the tool will employ some procedures, perhaps some AI techniques, to classify them as described in Chapter 4. This follows colour coding various types of resource links by applying the Web product. For example, the selectable region in Figure 4.2 could be indicated by enclosing the region inside a rectangular box or repainting some sections of the border region [2] as shown in Figure 13.1. Obviously, the dynamically changing nature of resources on the Web complicates the development of such a plug-in further. The feasibility of developing such a plug-in needs careful evaluation. Such software is unlikely to perform the work of a human Web page designer. This thesis has deliberately designed the Web colour scheme to make it is easy for the Web authors to apply it while designing the pages. If the colour design is adopted as a global convention by web designers, then a tool to recolour Web pages would not be necessary.

13.3.6 Software Agent

Development of an agent that would serve user needs and act on user behalf in terms of monitoring/managing user's selection of various Web resources and automatically colour coding the desirability and/or accessibility of content may be advantageous. The concerns

[1] During the course of the research, complexities associated with changing colour values of individual widgets (i.e. without repainting the whole screen) during runtime have directed the author to contact several experienced programmers including the author of "Editres". Most of them agreed with the difficulties identified in dynamically changing colours; some encouraged the author to go ahead with developing the colour filter along with a warning that it would require a considerable amount of time. Others described the tool as "not feasible" within the current limitations of existing software platforms.

[2] This approach may not be entirely suitable in cases where the visual integrity of the photographs/paintings need to be preserved while viewing them on the Web.

raised in the context of developing a possible plug-in (Section 13.3.5) also applies to the design and delivery of a software agent.

13.3.7 Other Under Exploited Features

In addition to colour, other features such as spatial location, apparent depth and motion also appear under exploited (see section 1.1). While the present work concentrates on the functionality of colour in WPD, a very similar approach could be taken to these features.

References

[9241-11 1997]
ISO/DIS 9241-11. *Ergonomic requirements for office work with visual display terminals (VDTs)– Part 8:Requirements for displayed colours.* Technical Report, International Organisation for Standardisation, Geneva, Switzerland, 1997.

[AldusCorp. 1991]
AldusCorp. *Aldus Supercard 1.6.* Seattle. WA: Aldus Corp., 1991.

[Allrich 1996]
Steve Allrich. *Oil Painting for the Serious Beginner: Basic Lessons in Becoming a Good painter.* Watson-Guptill Publications, July 1996.

[Almeida *et al.* 1995]
Ana Almeida, Licinio Roque, and Antonio Figueiredo. "Cyberspace: The HCI frontier? a new model in human-computer interaction". In Kirby M A R, Dix A J, and Finlay J E, editors, *Proceedings of the HCI'95 Conference: People and Computers X*, pages 51–62, Cambridge University Press, 1995.

[Alonso and Norman 1998]
Diane Lindwarm Alonso and Kent L Norman. "Apparency of contingencies in single panel and pull-down menus". *International Journal of Human-Computer Studies*, volume 49, number 1, pages 59–78, 1998.

[Altman 1991]
Altman. *Practical Statistics for Medical Research.* Chapman & Hall, 1991.

[Amir 1996]
D A Amir. *Fermat's Last Theorem: Unlocking the Secret of an Ancient Mathematical Problem.* Penguin Books, 1996.

[Anderson and Apperley 1990]
P S Anderson and M D Apperley. "An interface prototyping system based on lean cuisine". *Interacting with Computers*, volume 2, number 2, pages 217–226, 1990.

[Angel 1997]
Edward Angel. *Interactive Computer Graphics: A Top-Down Approach with OpenGL.* Addison-Wesley Pub Co (Net), 1997.

[Apperley and Spence 1989]
M D Apperley and R Spence. "Lean cuisine: A low fat notation for menus". *Interacting with Computers*, volume 1, number 1, pages 43–68, 1989.

[AppleComputerInc. 1991]
AppleComputerInc. *Hypercard 2.1*. Cupertino, CA: Apple Computer Inc., 1991.

[Baecker and Buxton 1987]
R M Baecker and W A S Buxton. *Readings in Human-Computer Interaction: A Multidisciplinary Approach*. Morgan kaufmann, 1987.

[Batchelor and Whelan 1993]
B G Batchelor and P F Whelan. "Generalisation procedures for colour recognition". In *Proceedings of the SPIE Conference on Machine Vision Applications, Architectures and System Integration II*, SPIE, 1993.

[Batchelor and Whelan 1997]
Bruce G Batchelor and Paul F Whelan. "Real-time colour recognition for machine vision systems". In *John Dalton's Colour Vision Legacy*, Taylor & Francis Ltd, 1997.

[Bellamy and Carroll 1990]
R K E Bellamy and J M Carroll. "Redesign by design". In D Diaper, D Gilmore, G Cockton, and B Shackel, editors, *Human-Computer Interaction: Interact' 90*, pages 199–205, North-Holland, 1990.

[Benest and Dukić 1993]
I D Benest and D Dukić. "Computer supported framework". In Diaper D and Sanger C, editors, *CSCW in Practice: An Introduction to Case Studies*, pages 127–150, Springer-Verlag: London, 1993.

[Benyon 1992]
D Benyon. "The role of task analysis in system design". *Interacting with Computers*, volume 4, number 1, pages 102–123, 1992.

[Benyon *et al.* 1990]
David Benyon, Laurie Keller, Jenny Preece, and Yvonne Rogers. *A Guide to Usability*. The Open University, 1990.

[Berlin and Kay 1999]
B Berlin and P Kay. *Basic Color Terms: Their Universality & Evolution*. Cambridge University Press, 1999.

[Bertin 1983]
Jacques Bertin. *Semiology of Graphics - translated by William J Berg*. University of Wisconsin Press, 1983.

[Bewley *et al.* 1983]
W L Bewley, T L Roberts, D Schroit, and W L Verplank. "Human factors testing in the

design of xerox's 8010 'star' office workstation". In *CHI'83 Conference on Human Factors in Computer Systems*, pages 72–77, ACM, 1983.

[Blake 1986]
T Blake. "Introduction to the art and science of user interface design". *Intutive Software and Interaction Systems, CA*, 1986.

[Blethyn and Parker 1990]
S G Blethyn and C Y Parker. *Designing Information Systems*. Butterworth-Heinemann, 1990.

[Boynton 1978]
R M Boynton. "Colour in contour and object perception". In Carterette and Friedman, editors, *Handbook of Perception*, pages 173–98, New York: Academic Press, 1978.

[Brake 1993]
I J Van Brake. "The plasticity of categories: The case of color". *British Journal of the Philosophy of Science*, volume XL 44, pages 103–35, 1993.

[Braudes 1991]
Robert E Braudes. "Conceptual modelling: A look at system-level user interface issues". In John Karat, editor, *Taking Software Design Seriously: Practical Techniques for Human-Computer Interaction*, chapter 10, pages 195–207, Academic Press, 1991.

[Braun *et al.* 1994]
Curt C Braun, Lori Sansing, Robert S Kennedy, and N Clayton Silver. "Signal word and color specifications for product warnings: An isoperformance application test and evaluation: Test and evaluation of product design [lecture]". In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, pages 1104–1108, 1994.

[Brown 1988]
C Brown. *Human-Computer Interface Design Guidelines*. Ablex, 1988.

[Brown and Cunningham 1989]
J R Brown and S Cunningham. *Programming the User Interface: Principles and Examples*. Wiley, 1989.

[Brown 1996]
Luanne Brown. "On the road to good web design". *http:// msdn.microsoft.com/ workshop/ design/ layout/ girvin.asp*, 1996.

[Cakir *et al.* 1980]
A Cakir, D J Hart, and T F M Stewart. *Visual Display Terminals: A Manual Covering Ergonomics, Workplace Design, Health and Safety, Task Organization*. New York: Wiley, 1980.

[Cameron 1983]
J R Cameron. *JSP & JSD: the Jackson Approach to Software Development*. IEEE Computer Society Press, 1983.

[Carroll 1991]
J M Carroll. "History and hysteresis in theories and frameworks for HCI". *People and Computers*, volume VI, pages 47–56, 1991.

[Carroll *et al.* 1988]
J M Carroll, R L Mack, and W A Kellogg. "Interface metaphors and user interface design". In *Handbook of Human-Computer Interaction*, Elsevier Science Publishers, B V, Amsterdam, 1988.

[Carter 1979]
R C Carter. "Search time with a color display: Analysis of distribution functions". *Human Factors*, volume 21, pages 293–302, 1979.

[Carter and Cahill 1979]
R C Carter and M C Cahill. "Regression models of search time for color-coded information displays". *Human Factors*, volume 21, pages 293–302, 1979.

[Carter and Crovella 2000]
R C Carter and M E Crovella. "Dynamic serve selection using bandwidth probing in wide-area networks". *http:// citeseer.nj.nec.com/ carter96dynamic.html*, 2000.

[Carver 1988]
M Carver. "Practical experience of specifying the human-computer interface using JSD". In Megaw E, editor, *Contemporary Ergonomics*, pages 177–182, London: Taylor and Francis, 1988.

[Chapanis 1965]
A Chapanis. "Color names for color space". *American Scientist*, volume 53, pages 327–346, 1965.

[Chase 1970]
W D Chase. "Evaluation of several TV display system configurations for visual simulation of the landing approach". *IEEE Transactions on Man-Machine Systems*, volume MMS-11, pages 140–149, 1970.

[Chen *et al.* 1990]
J R Chen, R K Belew, and G B Salomon. "A connectionist network for color selection". In *International Conference On Neural Networks*, 1990.

[Chevreul 1967]
M E Chevreul. *The Principles of Harmony and the Contrast of Colors*. New York: Reinhold. Originally published in 1854, 1967.

[Chin *et al.* 1988]
J P Chin, V A Diehl, and K L Norman. "Development of an instrument measuring user satisfaction of the human-computer interface". In *Human factors in computing systems*, pages 213–218, New York: Association of Computing Machinery, 1988.

[Christ 1975]
R E Christ. "Review and analysis of color coding research for visual displays". *Human Factors*, volume 17, number 6, pages 542–570, 1975.

[Christ and Corso 1983]
R E Christ and G M Corso. "The effects of extended practice on the evaluation of visual display codes". *Human Factors*, volume 25, pages 71–84, 1983.

[Christner and Ray 1961]
C A Christner and H W Ray. "An evaluation of the effect of selected combinations of target and background coding on map-reading performance: Exp v". *Human Factors*, volume 3, pages 131–146, 1961.

[Christopher D Wickens and Sarno 1995]
Melanie LaClair Christopher D Wickens and Kenneth Sarno. "Graph-task dependencies in three-dimensional data: Influence of three-dimensionality and color visual performance: Graphical user interfaces [lecture]". In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, pages 1420–1424, 1995.

[Christopher G Healey and Enns 1996]
Kellogg S Booth Christopher G Healey and James T Enns. "High-speed visual estimation using preattentive processing". *ACM Transactions on Computer-Human Interaction*, volume 3, number 2, pages 107–135, 1996.

[CNN 2001]
CNN. "Online news article: http:// www.cnn.com/ 1999/ tech/ computing/". *http:// www.cnn.com/ 1999/ TECH/ computing/ 12/ 09/ mcnealy.slams.gates.idg/*, 2001.

[Cockton 1990]
G Cockton. "Lean cuisine: No sauces, no courses!". *Interacting with Computers*, volume 2, number 2, pages 205–216, 1990.

[Collin 1989]
S M H Collin. *The Hamlyn Dictionary of Computing*. London: Hamlyn, 1989.

[Converse *et al.* 1992]
Sharolyn A Converse, Sandra Kozar, and David Batten. "Color coding to facilitate performance of focused attention tasks with object displays visual performance: Object and graphical displays". In *Proceedings of the Human Factors Society 36th Annual Meeting*, pages 1493–1497, 1992.

[Corbett and Kirakowski 1987]
M Corbett and J Kirakowski. "Computerizing data presentation and analysis". In Bullinger H and Shackel B, editors, *Interact'87: Human-Computer Interaction*, pages 879–884, North-Holland: Elsevier, 1987.

[Corte 1986]
W De Corte. "Finding appropriate colors for color displays". In *Colour Research and Application*, John Wiley & Sons, Inc., 1986.

[Crovella 2000]
M E Crovella. "Dynamic server selection in the internet". *http:// www.cs.bu.edu/ faculty/ crovella/ paper-archive/ hpcs95/ paper.html*, 2000.

[Cummaford and Long 1998]
S Cummaford and J Long. "Towards a conception of hci engineering design principles". In *Proceedings of the Ninth European Conference on Cognitive Ergonomics (ECCE9)*, pages 24–26, France:EACE, August 1998.

[Dalton 1997]
John Dalton. "John dalton's colour vision legacy". In *John Dalton's Colour Vision Legacy*, Taylor & Francis Ltd, 1997.

[Daniels 1979]
M I Daniels. *Statistics in Psychology: An Introductory Guide*. 1979.

[Davis 1990]
A M Davis. *Software Requirements: Analysis & Specification*. Prentice-Hall, 1990.

[Deitel *et al.* 2000]
Deitel, Deitel, and Nieto. *Internet & World Wide Web - How to Program*. Prentice Hall, 2000.

[Diaper 1989]
D Diaper. "Bridging the gulf between requirements and design". In *Simulation in the Development of User Interfaces*, pages 129–145, The Ergonomics Society Conference, 1989.

[Diaper 1982]
D Diaper. *Central Backward Masking and the Two-Task Paradigm*. PhD thesis, University of Cambridge, 1982.

[Diaper 1989]
D Diaper. "Designing expert systems: From dan to beershaw". In *Knowledge Elicitation: Principles, Techniques and Applications*, pages 15–46, Ellis Horwood, 1989.

[Diaper 1987]
D Diaper. "Designing systems for people: Beyond user centred design". In *Proceedings of the Share European Association (SEAS) Anniversary Meeting*, pages 283–302, 1987.

[Diaper 1989a]
D Diaper. "The discipline of human-computer interaction". *Interacting with Computers*, volume 1, number 1, pages 3–5, 1989.

[Diaper 1989b]
D Diaper. "Giving HCI away". In Sutcliffe A and Macaulay L, editors, *People and Computers*, pages 109–120, Cambridge: Cambridge University Press, 1989.

[Diaper 1988]
D Diaper. "Natural language communication with computers: Theory, needs and practice". In Duffin P, editor, *KBS in Government*, Blenheim Online, 1988.

[Diaper 1990]
D Diaper. "Simulation: A stepping stone between requirements and design". In Life A, Narborough-Hall C, and Hamilton W, editors, *Simulation and the User Interface*, pages 59–72, Taylor and Francis, 1990.

[Diaper 1989]
D Diaper. "Task observation for human-computer interaction". In Diaper D, editor, *Task Analysis for Human-Computer Interaction*, pages 210–237, Chinchester: Ellis Horwood, 1989.

[Diaper 1986]
D Diaper. "Will expert systems be safe?". In *Second International Expert Systems Conference*, pages 561–572, Oxford: Learned Information, 1986.

[Diaper and Addison 1992]
D Diaper and M Addison. "HCI: The search for solutions". In Monk A F, Diaper D, and Harrison M D, editors, *People and Computers*, pages 493–5, Cambridge University Press, 1992.

[Diaper and Addison 1992a]
D Diaper and M Addison. "Task analysis & systems analysis for software engineering". *Interacting with computers*, volume 4, number 1, pages 124–139, 1992a.

[Diaper and Addison 1991]
D Diaper and M Addison. "User modelling: The task oriented modelling (TOM) approach to the designer's model". In Diaper D and Hammond N, editors, *People and Computers VI*, pages 387–402, Cambridge University Press, 1991.

[Diaper and Sahithi 1995]
D Diaper and P S Sahithi. "Redfaces over user interfaces: What should colour be used for?". In Kirby M A R, Dix A J, and Finlay J E, editors, *People and Computers X: Proceedings of the HCI'95 Conference*, pages 425–435, September 1995.

[Dillon 1991]
A Dillon. "Human factors issues in the design of hypermedia interfaces". In Heather Brown, editor, *Hypermedia/ Hypertext and Object-oriented Databases*, pages 95–105, Chapman & Hall, 1991.

[Douglas and Kirkpatrick 1996]
Sarah Douglas and Ted Kirkpatrick. "Do color models really make a difference". In *Proceedings of ACM CHI 96 Conference on Human Factors in Computing Systems*, pages 399–405, 1996.

[Duijnhouwer and Zwaga 1995]
F Duijnhouwer and H J G Zwaga. "The influence of a fixed background in a vdu display on the efficiency of colour and shape coding computer displays i". In *Human Factors Perspectives on Human-Computer Interaction - Selections from the Human Factors & Ergonomics Society Annual Meetings 1983-1994*, pages 19–23, Santa Monica, California Human Factors and Ergonomics Society, 1995.

[Edwards and Hardman 1989]
D Edwards and L Hardman. "Lost in hyperspace: Cognitive mapping and navigation in a hypertext environment". In R McAleese, editor, *Hypertext: Theory into Practice*, Intellect, London, UK, 1989.

[Egeth and Pachella 1969]
H Egeth and R Pachella. "Multidimensional stimulus identification". *Perception & Psychophysics*, volume 5, pages 341–346, 1969.

[Elkins 1998]
James Elkins. *What Oil Painting is? How to Think About Oil Painting, Using the Language of Alchemy*. Amazon.com, November 1998.

[Evan 1995]
Thompson Evan. *Colour Vision*. London:Routledge, 1995.

[Faulkner 2000]
Xristine Faulkner. *Usability Engineering*. MacMillan Press Ltd, 2000.

[Festinger 1957]
L Festinger. *A Theory of Cognitive Dissonance*. Harper and Row, 1957.

[Foley *et al.* 1990]
J D Foley, A Van Dam, S K Feiner, and J F Hughes. *Computer Graphics: Principles and Practice*. The Systems Programming Series. New York: Addison-Wesley, 2 edition, 1990.

[Foley *et al.* 1997]
James D Foley, Andris Van Dam, Steven K Feiner, John F Hughes, and Richard L Phillips. *Introduction to Computer Graphics*. Addison-Wesley Publishing Company, Inc., 1997.

[Galitz 1985]
W O Galitz. *Handbook of Screen Format Design*. QED Information Sciences MA: Wellesley Hills, 2 edition, 1985.

[Galitz 1980]
W O Galitz. "Human factors in office automation". *Atlanta, GA: Life Office Management Association*, 1980.

[Garner 1972]
W R Garner. "Information integration and form of encoding". In Milton A W and Martin E, editors, *Coding Processes in Human Memory*, New York: John Wiley, 1972.

[Gasen 1995]
Jean B Gasen. "Support for HCI educators: A view from the trenches". In Kirby M A R, Dix A J, and Finlay J E, editors, *Proceedings of the HCI'95 Conference: People and Computers X*, pages 15–20, Cambridge University Press, 1995.

[Gerritsen 1988]
F Gerritsen. *Evolution in Color*. West Chester, PA: Schiffer Publishing Ltd. Originally published in Dutch in 1982 as Evolutie van de Kleurenleer. Translation by Dr Edward Force and Ruth de Vriendt, 1988.

[Gerritsen 1989]
F Gerritsen. *Theory and Practice of Color*. New York: Van Nostrand Reinhold Company. Translation of Het Fenomeen Kleur by Ruth De Vriendt, 1989.

[Gould and Lewis 1985]
J D Gould and C H Lewis. "Designing for usability: Key principles and what designers think". *Communications of ACM*, volume 28, number 3, pages 300–311, March 1985.

[Green *et al.* 1987]
Paul T Green, P E Miller-Jacobs, and Harold H Miller-Jacobs. "Color displays applied to command, control and communication (c3) systems???". In H John Durrett, editor, *Color and the Computer*, Academic Press, Inc., 1987.

[Green 1990]
T R G Green. "The cognitive dimension of viscosity: A sticky problem for HCI". In Diaper D, Gilmore D, Cockton G, and Shackel B, editors, *Human-Computer Interaction: Interact '90*, pages 70–86, North-Holland, 1990.

[Greene and d'Oliveira 1978]
Judith Greene and Manuela d'Oliveira. *Cognitive Psychology: Methodology handbook [Part 1]*. The Open University Press, 1978.

[Grudin 1989]
J Grudin. "The case against user interface consistency". *Communications of the ACM*, volume 32, number 10, pages 1164–1173, 1989.

[Grudin 1992]
J Grudin. "Consistency, standards, and formal approaches to interface development and evaluation - a overview". *ACM Transactions on Information Systems*, volume 10, number 1, pages 103–111, 1992.

[Grudin 1990]
J Grudin. "A foolish consistency". *BYTE*, volume 15, number 3, page 364, 1990.

[Gupta 1986]
S C Gupta. "Efficiency-consistency in designs". *Communications in Statistics-Theory and Methods*, volume 15, number 4, pages 1315–1318, 1986.

[Gygi 1990]
K Gygi. "Recognizing the symptoms of hypertext and what to do about it". In Laurel B, editor, *The Art of Human-Computer Interface Design*, Reading, MA: Addison-Wesley, 1990.

[Hammer 1984]
M Hammer. "The oa mirage". *Datamation*, pages 36–46, February 1984.

[Harper and Norman 1993]
B D Harper and K L Norman. "Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5". In *Mid Atlantic Human Factors Conference*, pages 224–228, February 1993.

[Haubner and Benz 1985]
P Haubner and C Benz. "Information display on monochrome and colour screens". pages 371–376, 1985.

[Hewitt and Gilbert 1993]
B Hewitt and G N Gilbert. "Groupware interfaces". In Diaper D and Sanger C, editors, *CSCW in Practice: An introduction and Case Studies*, pages 31–38, London: Springer-Verlag, 1993.

[Hill 1990]
F S Hill. *Computer Graphics*. Macmillan Publishing company, 1990.

[Hix and Hartson 1993]
D Hix and R Hartson. *Developing User Interfaces*. John Wiley, 1993.

[Hope and Walch 1990]
A Hope and M Walch. *The Color Compendium*. New York: Van Nostrand Reinhold, 1990.

[Hubbard 1992]
R E Hubbard. "Molecular graphics: From pen plotter to virtual reality". In Monk A, Diaper D, and Harrison M D, editors, *People and Computers*, pages 21–8, Cambridge University Press: Cambridge, 1992.

[Hunt 1994]
R W G Hunt. *Foreword: Computer generated color - A prctical guide to presentation and display*. John Wiley & Sons Ltd, 1994.

[Hunt 1991]
R W G Hunt. *Measuring Colour*. Ellis Horwood, 2 edition, 1991.

[Hunt 1995]
R W G Hunt. *The Reproduction of Colour*. Fountain Press, 5 edition, 1995.

[Ishihara 1998]
S Ishihara. *Ishihara's Tests for Colour Deficiency.* Amazon, 1998.

[Jacko *et al.* 1999]
Julie A Jacko, Max A Dixon, Robert H Rosa, Ingrid U Scott, and Charles J Pappas. "Visual profiles: A critical component of universal access profiles, notes, and surfaces". In *Proceedings of ACM CHI 99 Conference on Human Factors in Computing Systems*, pages 330–337, 1999.

[Jackson *et al.* 1994]
Richard Jackson, Lindsay MacDonald, and Ken Freeman. *Computer generated color - A prctical guide to presentation and display.* John Wiley & Sons Ltd, 1994.

[Jadad 2001]
Alejandro R Jadad. "Randomised controlled trials - a user's guide". *http:// www.bmjpg.com/ rct/ chapter1.html*, 2001.

[Jeffrey and Beck 1972]
T E Jeffrey and F J Beck. *Intelligence Information from Total Optical Color Imaginery.* Technical Report, U S Army Behaviour and Systems Research Laboratory, Research Memorandum, November 1972.

[Johnson 1995]
Chris Johnson. "Time and the web: Representing and reasoning about temporal properties of interaction with distributed systems". In Kirby M A R, Dix A J, and Finlay J E, editors, *Proceedings of the HCI'95 Conference: People and Computers X*, pages 39–50, Cambridge University Press, 1995.

[Johnson and Solso 1978]
Homer H Johnson and Robert L Solso. *An introduction to experimental design in Psychology: A case approach.* Harper & Row, Publishers, Inc, second edition, 1978.

[Johnson and Roberts 1989]
Jeff Johnson and Teresa L Roberts. "The xerox star: A retrospective". *IEEE Spectrum*, volume 9, number 11, pages 11–28, October 1989.

[Kaipa 1998]
Prasad Kaipa. "Designing organizations that learn: An executive guide to learning". *www.mithya.com/ learning/ designorg.html; First published in Chinmaya Management Review*, 1998.

[Kay and the Learning Research Group 1976]
A Kay and the Learning Research Group. *Personal dynamic media.* Technical Report, Xerox Palo Alto Research Center Technical Report SSL-76-1, (see alo IEEE Computing, March 1977, 31-41), 1976.

[Kellogg 1987]
W A Kellogg. "Conceptual consistency in user interface : Effects on user performance".

In *Proceedings of INTERACT'87 Conference of Human-Computer interaction*, Stuttgart, September 1987.

[Kieras 1988]
D E Kieras. *Towards a practical GOMS model methodology for user interface design*, pages 135–157. North-Holland: Amsterdam, 1988.

[Kopala 1979]
Carole J Kopala. "The use of color-coded symbols in a highly dense situation display". *Proceedings of Human factors Soc.*, number 23, pages 397–401, 1979.

[Koritzinsky 1989]
Ianne Howards Koritzinsky. *New ways to consistent interfaces*, chapter 8, pages 93–106. Academic Press, Inc., 1989.

[Kraupner-Stadler 1991]
H Chr Kraupner-Stadler. "Fundamentals for the use of colors in user interfaces congress i: Work with terminals: Input and display devices; display ergonomics". In *Proceedings of the Fourth International Conference on Human-Computer Interaction*, pages 98–102, 1991.

[Laar *et al.* 1997]
Darren Van Laar, Richard Flavell, Ian Umbers, and Jonathan Smalley. "A methodology for producing maximally discriminable, nameable colours in control room displays". In *John Dalton's Colour Vision Legacy*, Taylor & Francis Ltd, 1997.

[Lalomia and Happ 1987]
Mary J Lalomia and Alan J Happ. "The effective use of color for text on the ibm 5153 color display random access ii". In *Proceedings of the Human Factors Society 31st Annual Meeting*, pages 1091–1095, 1987.

[Landauer 1993]
Thomas Landauer. *The Trouble with Computers*. MIT Press, 1993.

[Landow 1989]
G P Landow. "The rhetoric of hypermedia: Some rules for authors". *Journal of Computing in Higher Education*, volume 1, number 1, 1989.

[Lang *et al.* 1994]
Virginia A. Lang, Michael Keith, and Andrew Kavie. "Design characteristics for foveal and peripheral tasks in multi-task visual displays special sessions: Posters". In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, page 956, 1994.

[Lansdale and Ormerod 1994]
M W Lansdale and T C Ormerod. *Understanding Interfaces: A hand book of human-computer dialogue*. Academic press, 1994.

[Lansdale *et al.* 1989]
M W Lansdale, D R Young, and C A Bass. "MEMORIES: A personal multimedia information

system". In Sutcliffe A and Macaulay L, editors, *People and Computers*, pages 315–327, Cambridge University Press, 1989.

[Lee *et al.* 1994]
A Y Lee, P W Foltz, and P G Polson. "Memory for task-action mappings - mnemonics, regularity and consistency". *International Journal of Human-Computer Studies*, volume 40, number 5, pages 771–794, 1994.

[Lee and Fisk 1993]
M D Lee and A D Fisk. "Disruption and maintenance of skilled visual - search as a function of degree of consistency". *Human Factors*, volume 35, number 2, pages 205–220, 1993.

[Legge 1989]
G E Legge. "Reading: Effects of contrast and spatial frequency". *Applied Vision: Technical Digest Series*, volume 16, pages 90–93, 1989.

[Licklider 1979]
J C R Licklider. "Impact of information technology on education in science and technology". In *Technology in Science Foundation, the next ten years: Perspectives and Recommendations*, 1979.

[Long 1986a]
J Long. "People and computers: Designing for usability". In M D Harrison and A F Monk, editors, *Proceedings of the Second Conference of the British Computer Society*, pages 3–22, University of York, Cambridge University Press, September 1986.

[Long 1986b]
J Long. "People and computers: Designing for usability". In Harrison M and Monk A, editors, *People and Computers: Designing for Usability*, pages 3–23, Cambridge University Press: Cambridge, U.K., 1986.

[Long and Dowell 1989a]
J Long and J Dowell. "Conceptions of the discipline of hci: Craft, applied science, and engineering". In *Proceedings of the fifth conference*, pages 9–32, University of Nottingham, Cambridge University Press, September 1989.

[Long and Dowell 1989b]
J Long and J Dowell. "Conceptions of the discipline of HCI: Craft, applied science, and engineering". In *People and Computers V*, pages 9–34, Cambridge University, 1989.

[Löwgren and Laurén 1993]
J Löwgren and U Laurén. "Supporting the use of guidelines and style guides in professional user interface design". *Interacting with Computers*, volume 5, number 4, pages 385–396, 1993.

[Lynch 1994a]
Patrick J Lynch. "Visual design for the user interface part 1: Design fundamentals". *Journal of Biocommunications*, volume 21, number 1, pages 22–30, 1994.

[Lynch 1994b]
Patrick J Lynch. "Visual design for the user interface part 2: Graphics in the interface". *Journal of Biocommunications*, volume 21, number 2, pages 6–15, 1994.

[Lynch and Horton 1997]
Patrick J Lynch and S Horton. "Yale style manual". *http://info.med.yale.edu/caim/manual/contents.html*, 1997.

[Lynch and Jaffe 1990]
Patrick J Lynch and C J Jaffe. "An introduction to interactive hypermedia". *Journal of Biocommunications*, volume 17, number 1, pages 2–8, 1990.

[MacDonald 1999]
Lindsay W MacDonald. "Tutorial: Using color effectively in computer graphics". *IEEE Computer Graphics and Applications*, volume 19, number 4, pages 20–35, 1999.

[MacDonald and Vince 1994]
Lindsay W MacDonald and John Vince. *Interacting with Virtual Environments*. John Wiley, 1994.

[Macintyre 1991]
Blair Macintyre. *A Constraint-based Approach to Dynamic Colour Management for Windowing Interfaces*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 1991.

[Macworth 1963]
J F Macworth. "The relation between the visual image and post-perceptual immediate memory". *Journal of Verbal Learning and Verbal Behaviour*, volume 2, pages 75–85, 1963.

[Markoff 1972]
J I Markoff. *Traget Recognition Performance with Chromatic and Achromatic Displays*. Volume SRM-148, Honeywell, Inc., January 1972.

[Maund 1998]
Barry Maund. "Stanford encyclopedia of philosophy". *http://plato.stanford.edu/archives/spr1998/entries/color*, 1998.

[McManus 1995]
Chris McManus. *Foreword: The beginners guide to colour psychology*. Kyle Cathie Limited, 1995.

[Meier 1988]
B J Meier. "Ace: A color expert system for user interface design". In *ACM SIGGRAPH Symposium on User Interface Software*, pages 117–128, 1988.

[Meier 1987]
B J Meier. *Effective Use of Color in User-Computer Interface Design*. Technical Report, Brown University, Department of Computer Science, Box 1910, Providence, RI 02912, 1987.

[Miller 1975]
S Miller. *Experimental Design and Statistics*. Methuen, 1975.

[Mollon 1982]
J D Mollon. *The senses*. Cambridge University Press, 1982.

[Munns 1968]
M Munns. "Some effects of display symbol variation upon operator performance in aircraft interception". *Perception and Motor Skills*, number 26, pages 1215–1221, 1968.

[Munsell 1947]
A H Munsell. *A Color Notation: An Illustrated System Defining All Colors and Their Relations be Measured Scales of Hue, Value and Chrome*. Munsell Color Company, Baltimore, MD, 1947.

[Neale 1993]
R G Neale. "Consistency and concerns". *Electrical Engineering*, volume 65, number 5, page 800, 1993.

[Newell 1984]
A Newell. "Speech: The natural modality for man-machine interaction?". In Shackel B, editor, *Interact'84 - First IFIP conference on Human-Computer Interaction*, pages 174–178, North Holland: Elsevier, 1984.

[Newman and Lamming 1995]
William Newman and Michael G Lamming. *Interactive System Design*. Addison-Wesley, 1995.

[Niederst 1999]
Jennifer Niederst. *Web Design In a Nutshell*. O'Reilly & Associates, Inc, 1999.

[Nielsen 1989]
Jakob Nielsen. "Coordinating user interfaces for consistency". *SIGCHI Bulletin*, volume 20, number 3, pages 63–65, January 1989.

[Nielsen 2000]
Jakob Nielsen. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, 2000.

[Nielsen 1989]
Jakob Nielsen. *Executive summary: Coordinating user interfaces for consistency*, chapter 1, pages 1–8. Academic Press, Inc., 1989.

[Nielsen 1998]
Jakob Nielsen. "Sun's new web design". *http:// www.sun.com/ 980113/ sunonnet/*, January 1998.

[Nielsen 1993]
Jakob Nielsen. *Usability Engineering*. Academic Press, Inc.; Harcourt Brace & Company, Publishers, 1993.

[Nijjar 1976]
L J Nijjar. *Using color effectively (or peacocks can't fly)*. Technical Report, IBM Corporation(IBM TR52.0018), Atlanta, GA, 1976.

[Norman 1990a]
D A Norman. "Why interfaces don't work". In Laurel B, editor, *The art of Human-Computer Interface Design*, Reading, MA: Addison-Wesley, 1990.

[Norman 1990b]
Ket L Norman. "Quis". *http:// lap. umd.edu/ QUISFolder/ newspaper1.html*, 1990.

[Nugent 1994]
William A Nugent. "Effects of symbol type, selection tool, and information density on tactical display visual search performance visual performance: Visual search performance [lecture]". In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, pages 1256–1260, 1994.

[Ostwald 1931]
W Ostwald. *Color Science: A Handbook of Advanced Students in Schools, Colleges, and the Various Arts, Crafts, and Industries Depending on the Use of Color*. Winsor and Newton, 1931.

[Pangalos 1992]
G J Pangalos. "Consistency and standardization of user interfaces". *Information and Software Technology*, volume 34, number 6, pages 397–401, 1992.

[Pantone 1999]
Pantone. "Colourweb". *www.pantone.com/ allaboutcolor/*, 1999.

[Perlman 1989]
Gary Perlman. *Coordinating Consistency of user interfaces, code, online help, and documentation with multilingual/ multitarget software specification*, chapter 4, pages 35–56. Academic Press, Inc., 1989.

[Pfleeger 1991]
S L Pfleeger. *Software Engineering: The Production of Quality Software*. Macmillan, 2 edition, 1991.

[Piaget 1970]
J Piaget. "Piaget's theory". In *Carmichael's Manual of Child Psychology*, New York: John Wiley, 1970.

[Porter 1967]
Lambert Cedric Porter. *Taxonomy of Flowering Plants*. Freeman, San Francisco, 1967.

[Prieditis 1988]
A Prieditis. *Analogica*. London: Pitman; California: Morgan Kaufmann, 1988.

[Priester 2001]
Gary W Priester. "All you need to know about web safe colors". *http:// webdevelopersjournal.com/ articles/ websafe1/*, 2001.

[Quiller 1989]
Stephen Quiller. *Color Choices*. Watson-Guptill Publications, New York, 1989.

[Ravden *et al.* 1989]
C Ravden, R Lewis, and R D Cooper. *Evaluating Usability of Human-Computer Interfaces: A Practical Method*. Ellis Horwood, 1989.

[Reising and Aretz 1987]
John M Reising and Anthony J Aretz. "Color computer graphics in military cockpits". In H John Durrett, editor, *Color and the Computer*, Academic Press, Inc., 1987.

[Ringel and Hammer 1964]
S Ringel and C Hammer. *Information Assimilation from Alphanumeric Displays: Amount and Density of Information Presented*. Technical Report TRN141, U S Army Personnel Research Office, Washington, DC, 1964.

[Rivlin *et al.* 1990]
C Rivlin, R Lewis, and R D Cooper. *Guidelines for screen design*. Blackwell, 1990.

[Robertson 1979]
P J Robertson. *A Guide to Using Colour on Alphanumeric Displays*. Unrestricted, IBM United Kingdom Laboratories Limited, December 1979.

[Robertson 1982]
P J Robertson. *Review of Colour Display Benefits*. Technical Report, Report of the IBM Hursley Human Factors Laboratory (HF 056), Hursley Park, UK, 1982.

[Robertson 1976]
P J Robertson. *The Use of Colour for Computer Displays*. Technical Report, Report of the IBM Hursley Human factors Laboratory (HF 005), Hursley Park, UK, 1976.

[Rosenberg 1988]
D Rosenberg. "Corporate user interface standards: A new component of product form". *Innovation*, volume 7, number 2, pages 25–28, 1988.

[Rosenberg 1989]
D Rosenberg. *A cost benefit analysis for corporate user interface standards: What price to pay for a consistent "Look and Feel"?*, chapter 3, pages 21–34. Academic Press, Inc., 1989.

[Sahithi and Diaper 1998]
P S Sahithi and D Diaper. "A style guide for background/ border colour in web page design". In *CIM'98 - Colour Imaging in Multimedia*, pages 287–295, University of Derby, U K, 1998.

[Sands 1963]
Feynman Leighton Sands. *The Feynman Lectures on Physics*. Addison-Wesley,, 1963.

[Sankar *et al.* 1995]
C S Sankar, F N Ford, and M Bauer. "A DSS user-interface model to provide consistency and adaptability". *Decision Support Systems*, volume 13, number 1, pages 93–104, 1995.

[Sassie 1997]
Martina Angels Sassie. *Eliciting and Describing Users' Models of Computer Systems*. PhD thesis, The University of Birmingham, UK, April 1997.

[Saunders *et al.* 1998]
D Saunders, J Cupitt, R Pillay, and K Martinez. "Maintaining accurate colour in images transferred across the internet". In *CIM 98: Colour Imaging in Multimedia*, 1998.

[Schutz 1961]
H G Schutz. "An evaluation of methods for presentation of graphic multiple trends. exp iii". *Human Factors*, number 3, pages 108–119, 1961.

[Shackel 1986]
B Shackel. "Ergonomics in design for usability". In Harrison M D and Monk A F, editors, *People and Computers: Designing for Usability. Proceedings of the Second Conference of the BCS HCI Specialist Group*, Cambridge University Press, 1986.

[Shepherd 1989]
A Shepherd. "Analysis and training in information technology tasks". In Diaper D, editor, *Task Analysis for Human-Computer Interaction*, pages 15–55, Ellis Horwood:Chinchester, 1989.

[Shneiderman 1994]
Ben Shneiderman. "Book preview of interactions". *Interactions*, pages 67–71, January 1994.

[Shneiderman 1992]
Ben Shneiderman. *Designing the User interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley publishing company, 2 edition, 1992.

[Shneiderman 1987]
Ben Shneiderman. *Designing the User interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley publishing company, 1 edition, 1987.

[Shneiderman 1997]
Ben Shneiderman. *Designing the User interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley publishing company, 3 edition, 1997.

[Shneiderman 1983]
Ben Shneiderman. "Direct manipulation: A step beyond programming languages". *IEEE Computer*, pages 57–68, August 1983.

[Shneiderman 1998]
Ben Shneiderman. "Quis: Online report". *http:// lap.umd.edu /*, 1998.

[Sidorsky *et al.* 1984]
R C Sidorsky, R N Parrish, J L Gates, and S J Munger. *Design guidelines for user transactions with battlefield automated systems:Prototype for a handbook.* ARI Research Product 84-08. Alexandria, VA, 1984.

[Silverstein 1987]
Louis D Silverstein. "Human factors for colour display systems: concepts, methods, and research". In H John Durrett, editor, *Color and the Computer*, Academic Press, Inc., 1987.

[Simpson 1961]
George Gaylord Simpson. *Principles of Animal Taxonomy*. New York: Columbia University Press, 1961.

[Simpson 1999]
Nichole Simpson. "Managing the use of style guides in an organisational setting: Practical lessons in ensuring ui consistency". *Interacting with Computers*, volume 11, number 3, pages 323–351, 1999.

[Slaughter *et al.* 1995]
Laura Slaughter, Kent L Norman, and Ben Shneiderman. "Assessing users' subjective satisfaction with the information system for youth services (ISYS)". number CS-TR-3463, March 1995.

[Smallman and Boynton 1994]
H S Smallman and R M Boynton. "Segragation of basic colors in an information display". *Journal of Optical Society of America*, volume 7, number 10, 1994.

[Smith 1978]
A R Smith. "Color gamut transform pairs". *SIGGRAPH*, pages 12–19, 1978.

[Smith *et al.* 1982]
D C Smith, C Irby, R Kimbell, B Verplank, and E Harslem. "Designing the star user interface". *BYTE*, volume 7, number 4, pages 242–282, 1982.

[Smith *et al.* 1983]
D C Smith, C Irby, R Kimbell, W Verplank, and E Harslem. "Designing the star user interface". In Degano P and Sandewall E, editors, *Integrated Interactive Computing Systems*, pages 297–313, Amsterdam: North-Holland Publishing Company, 1983.

[Smith 1987]
R Smith. *Programming the User Interface*. Wiley, 1987.

[Smith and Mosier 1986]
S L Smith and J N Mosier. *Guidelines for Designing User Interface Software*. Technical Report, Tech.Rept. ESD-TR-86-278, MITRE, Bedford, Mass.,, 1986.

[Smith *et al.* 1995]
Walter Smith, John Dunn, Kim Kirsner, and Mark Randell. "Colour in map displays: Issues for task-specific display design". *Interacting with computers*, volume 7, number 2, pages 151–165, 1995.

[Spool 1997]
Jared M Spool. *Web Site Usability: A Designer's Guide*. User Interface Engineering; North Andover, Massachusetts, 1997.

[Staffordshire University 1998]
UK: BCSHCI Staffordshire University. "The active web". *British HCI Mailing List*, December 1998.

[Sticker 1992]
Markus A Sticker. *Color and Geometry as Cues for Indexing*. Technical Report, NCSRTL, Chicago, http:// lite.ncstrl.org:3803/ Dienst/ UI/ 2.0/ Describe, November 1992.

[Stokes *et al.* 1996]
Michael Stokes, Matthew Anderson, Srinivasan Chandrasekar, and Ricardo Motta. "A standard default color space for the internet". *http:// www.w3.org/ Graphics/ Color/ sRGB*, 1996.

[Sullivan 1997]
John Sullivan. *http:// www.raritanval.edu/ departments/ cis/ Colorguide.html*, 1997.

[Swan and Perlman 1995]
J Edward Swan and Gary Perlman. "Color versus texture coding to improve visual search performance computer systems: Visual display of information". In *Human Factors Perspectives on Human-Computer Interaction - Selections from the Human Factors & Ergonomics Society Annual Meetings 1983-1994*, pages 321–325, Santa Monica, California Human Factors and Ergonomics Society, 1995.

[Tanaka *et al.* 1991]
Toshiki Tanaka, Ray E Eberts, and Gavriel Salvendy. "Consistency of human-computer interface design: Quantification and validation". *Human Factors*, volume 33, number 6, pages 653–676, 1991.

[Tero and Briggs 1994]
A Tero and P Briggs. "Consistency versus compatibility - a question of levels". *International Journal of Human-Computer Studies*, volume 40, number 5, pages 879–894, 1994.

[Tero and Brigg 1994]
Alan Tero and Pamela Brigg. "Consistency versus compatibility: a question of levels?". *International Journal of Human-Computer Studies*, volume 40, pages 879–893, January 1994.

[Thorndike 1913]
E L Thorndike. *Education Psychology: The Psychology of Learning*. New York: Teachers College Press, 1913.

[Tognazzini 1989]
Bruce Tognazzini. *Achieving consistency for the Macintosh*, chapter 5, pages 57–74. Academic Press, Inc, 1989.

[Travis 1990]
D Travis. "Applying visual psychphysics to user interface design". *Behaviour and Information Technology*, volume 9, number 5, pages 425–438, 1990.

[Travis 1999]
David Travis. "Article on colour". *http:// systemconcepts.co.uk / c1.htm*, 1999.

[Travis 1991]
David Travis. *Effective Color Displays: Theory and Practice*. Academic Press, 1991.

[Treisman 1982]
A Treisman. "A perceptual grouping and attention in visual search for features and for objects". *Journal of Experimental Psychology: Human Perception and Performance*, volume 8, pages 194–214, 1982.

[Treisman and Gelade 1980]
A Treisman and G Gelade. "A feature-integration theory of attention". *Canadian Journal of Psychology*, number 12, pages 97–136, 1980.

[Treisman and Gormican 1988]
A Treisman and Gormican. "Feature analysis in early vision: Evidence from search asymmetries". *Psychological Review*, number 95, pages 15–48, 1988.

[Tufte 1990]
Edward R Tufte. *Envisioning Information*. Graphics, Cheshire (Conn.), 1990.

[Tullis 1981]
Thomas S Tullis. "An evaluation of alphanumeric, graphic and color information displays". *Human Factors*, number 23, pages 541–550, 1981.

[Tullis 1983]
Thomas S Tullis. "The formatting of alphanumeric displays: A review and analysis". *Human Factors*, volume 25, number 6, pages 657–682, 1983.

[Tullis 1984]
Thomas S Tullis. *Predicting the Usability of Alphanumeric Displays*. PhD thesis, Rice University, Lawrence, kankas, 1984.

[Tullis 1988]
Thomas S Tullis. "Screen design". In Hellander M, editor, *Handbook of Human-Computer Interaction*, pages 377–411, Elsevier, 1988.

[URL 1998a]
URL. *http:// cs-tr.cs.berkeley.edu /*, 1998.

[URL 1998b]
URL. *http:// bulldog.prn.bc.ca/ TheSciences/ .. / ClrBlind/*, 1998.

[URL 2001a]
URL. "Color and animation: Kde user interface guidelines". *http:// developer.kde.org/ documentation/ design/ ui/ colour.html*, 2001.

[URL 2001b]
URL. "Color principles". *http:// www2.ncsu.edu:8010/ scivis/ lessons/ colormodels/*, 2001.

[URL 2000]
URL. "Example graphics". *www.sigmaxi.org/ amsci/ amsci.html & www.mcs.surrey.ac.uk/ Personal/ R.Knott/*, 2000.

[URL 2001]
URL. "Graphical user interface timeline". *http:// www.pla-netx.com/ linebackn/ guis /gui-timeline.htm*, February 2001.

[URL 1998]
URL. "Internet domain survey". *http:// www.nw.com/ zone/ www/*, November 1998.

[URL 2000a]
URL. "Web performance summary: Optimal application vantage". *www.optimal.com*, 2000.

[URL 2000b]
URL. "Web proxy severs". *vms.process.com*, 2000.

[URL 1998]
URL. "Worldwide domain statistics". *http://www.domainstats.com/*, November 1998.

[Vanniamparampil *et al.* 1995]
A Vanniamparampil, Ben Shneiderman, C Plaisant, and A Rose. *User Interface Re-engineering: A Diagnostic Approach*. Technical Report, Department of Computer science, University of Maryland, 1995.

[Verplank 1986]
W Verplank. *Designing Graphical User Interfaces*. ID TWO, CA, 1986.

[Waters and Mundy 1996]
Crystal Waters and Andrew Mundy. *Web concept & design: A comprehensive guide for creating effective Web sites*. New Riders Publications, 1996.

[Weinman and Heavin 1997a]
Lynda Weinman and Bruce Heavin. *Coloring Web Graphics 2*. New Riders Publishing, 1997.

[Weinman and Heavin 1997b]
Lynda Weinman and Bruce Heavin. *<Coloring Web Graphics.2> The Definite Resource for on the Web*. New Riders Publishing, 1997.

[Whiteside *et al.* 1988]
J A Whiteside, J L Bennett, and K A Holtzblatt. *Usability Engineering: Our Experience and Evolution*, pages 791–817. Amsterdam: North-Holland, 1988.

[Wiecha 1992]
Charles Wiecha. "ITS and user interface consistency - a response". *ACM Transactions on Information Systems*, volume 10, number 1, pages 112–114, 1992.

[Wolf 1989]
Richard Wolf. *Consistency as Process*, chapter 7, pages 89–92. Academic Press, Inc., 1989.

[Wright 1995]
Angela Wright. *The beginners guide to colour psychology*. Kyle Cathie Limited, 1995.

[Wyszecki 1982]
G Wyszecki. "A regular rhombohedral lattice sampling of munsell rotation space". *Journal of Optical Society of America*, volume 44, 1982.

[Wyszecki and Stiles 1982]
G Wyszecki and W Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. New York: Wiley, 1982.

Appendix A

User Interface Issues

A.1 HCI Issues

A.1.1 Prototyping in Redesign

There is a personal as well as cultural inertia to interface design which makes users content with an interface that is similar to their previous computer experience. Icons, for example, may be chosen for their visual appearance based on what experienced users are already familiar with such that experimental evaluation of icon design is unnecessary, in contrast to Bewley *et al.*'s [Bewley *et al.* 1983] work on the Star icon set. These days users expect document, folder, file, waste basket/trash can, etc. icons to have a style of appearance, even if different systems represent the icons slightly different. Overall, it can be concluded that evaluating icon sets may only be necessary where the function of the icon, appearance, its context, or all are likely to be novel to most users.

The inertia, either at individual or organisational level, or the cultural level, causes Carroll [Carroll 1991] to be critical of the rapid prototyping, "evolutionary", "participative design" approach which he describes as "clearly ascendant in current HCI". At the least, not all user perseveration between systems will be desirable yet the prototyping approach encourages this behaviour. As Carroll says of users involved in such a design approach:

> "... people who invest the effort to learn a system can come to see the system more positively than is warranted [Festinger 1957]. Perhaps people who take the trouble to learn a considerable set of systems, providing feedback on each in turn, might ineluctably come to see the final version as fabulously useful and usable."

Indeed, Carroll argues for a "contextualist critique" and he suggests:

> "... that HCI is fundamentally historical: to have an adequate practical understanding of a situation of use, one must understand more than the proximal context (the situation itself and the situation a moment ago), one must understand *how the situation was caused historically*[1]."

Bellamy and Carroll [Bellamy and Carroll 1990] on "Redesign by Design" studied how designers working with Smalltalk/V tackled the redesign of system tools. They conclude that:

> "We found that designers do consider usability issues, but sometimes not *all*[2] the usability strengths and weaknesses ... We show how psychological claims analysis can support and guide such redesign work."

In summary, against the mainstream approach in HCI today, Carroll argues and demonstrates that there remains a utility in HCI to a theoretical and psychological alternative approach which could augment a process of design based on redesign. What remains at issue may be the extent to which theoretical approaches can supplement the current, basically atheoretical, participative rapid prototyping approach.

A.1.2 Frequency and Safety

Kay's maxim [Kay and the Learning Research Group 1976, Smith *et al.* 1983] states that:

Simple things should be simple; complex things should be possible.

In many cases, what users want to do, and what the computer system is able to do, however are genuinely complicated. Indeed, it may well be the case that users *rarely* wish to do many simple things with a computer and thus a better maxim, stressing frequency, might be:

Frequent things should be simple.

Sophisticated interfaces, in theory at least, can reduce user errors and/or the time it takes to do something. Maximum benefit for the interface design effort, with one proviso, should be related to the frequency and duration of an operation; one example of a frequent operation is simply selecting a menu. After all, reducing the time it takes to do an

[1] Carroll's emphasis

[2] Their emphasis

infrequent operation (e.g. starting a nuclear station), even by an order of magnitude, will not make much overall difference, unless it also increases the frequency of the operation.

The proviso, however, concerns safety. While safety is recognised in "safety critical systems" such as air traffic control, nuclear power plants etc., most of which involve real time planning and control, the concept of safety in HCI has not been as widely employed in other application areas. Diaper [Diaper 1986, Diaper 1989a] suggested that it is necessary to have a broad view of safety which should encompass not only individuals who use a computer system but also "indirect end users" [Diaper 1987] who do not use the systems but are affected by them (e.g. colleagues); organisations (e.g. commercial companies); and larger groups (e.g. socio-economic classes). Any concept of safety must also include some notion of the consequences of a failure, which can range from the death of many people to minor, immediate frustration. No matter how discredited Hierarchial Task Analysis' "stopping rule" [Shepherd 1989] now is, it at least encapsulated the twin concepts of frequency (P: Probability of inadequate performance) and consequence (C: Cost of inadequate performance). A better maxim for user interface design might be:

Frequent things should be simple; all things should be as safe as possible.

Of course, the extent of possible safety includes interacting with other parameters, for example the costs involved.

A.1.3 Metaphors

Prieditis [Prieditis 1988] comments that it is peculiar for a widely used metaphor such as STAR's 'desktop' not to be subjected to more rigorous and formal scrutiny by the HCI community. Informally, and described in more detail by Smith *et al.* [Smith *et al.* 1983], the problems of analogy and metaphor involve scope and mapping relationships. Basically the idea is that there is some element in the real world which has an "is like" relationship to the thing modelled, for example a real desktop is like Star's screen. The problems with such relationships include: (1) there are elements in the real world that do not map to the metaphor, and *vice versa*, (2) there may be elements in the real world that map incorrectly to the metaphor, and *vice versa*, and so on.

To illustrate how the use of metaphor can cause usability problems, Benest and Dukić [Benest and Dukić 1993] use the "Automated Office Metaphor", as an extension of the desktop metaphor, of a book shelf to display available files and directories to users. The user selects a book to access the file or directory. The display draws both the book shelf itself and artistic book-ends, which of course are elements on the screen that are not selectable. Selecting a book does not remove it from the shelf but merely high-lights it as selected. The books are all of the same size, style, etc. on the shelf so that the powerful visual cues available on real book shelves are not available. Finally, the use of this metaphor introduces a major usability disadvantage because the "titles" are printed sideways (i.e. bottom to top) on the electronic book spines which, of course, makes them hard to read. It is difficult to believe that any advantage this metaphor provides, in terms of users being able to intuit functions, is more than out weighed by its usability disadvantages.

In conclusion, while there may be some mileage in using metaphors it is necessary to understand their scope and limitations at the design stage and not to use them merely because metaphors have become a "style". The extension of the desktop metaphor to rooms, libraries and so forth needs very careful consideration which does not appear to have been given in much recent HCI work.

A.2 Examples

Figure A.1 summarises the history of Star development until 1989, showing how various systems related to Star have influenced the others over the years. Double arrows indicate direct successors (i.e. follow-on versions). Many "influence arrows" are due to key designers changing jobs or applying concepts from their graduate research to products [Johnson and Roberts 1989].

Since then the timeline to date includes several other GUI based systems, for example Windows 3.0 - 1990, Amiga workbench - 1992, Windows 3.1 - 1992, Photon microGUI - 1994, Windows 95 - 1995, Windows NT 4.0 - 1996, Mac OS8 - 1997, Windows 98 - 1998, 98 Lite - 1998, RISC OS4 - 1999, Mac OS X - 1999, Windows 2000 - Aka Windows NT5 [URL 2001].

A.3 Under Exploited Display Features

A.3.1 Spatial location

Tullis [Tullis 1988] reports his work [Tullis 1981] that "redesigning a key display from a system for testing telephone lines resulted in a 40% reduction in the time required by the users to interpret the display" which lead to a saving of 79 person years per year. Similarly Tullis [Tullis 1984] reports a difference of 128% between the best and worst format displays for information extraction from an airline system.

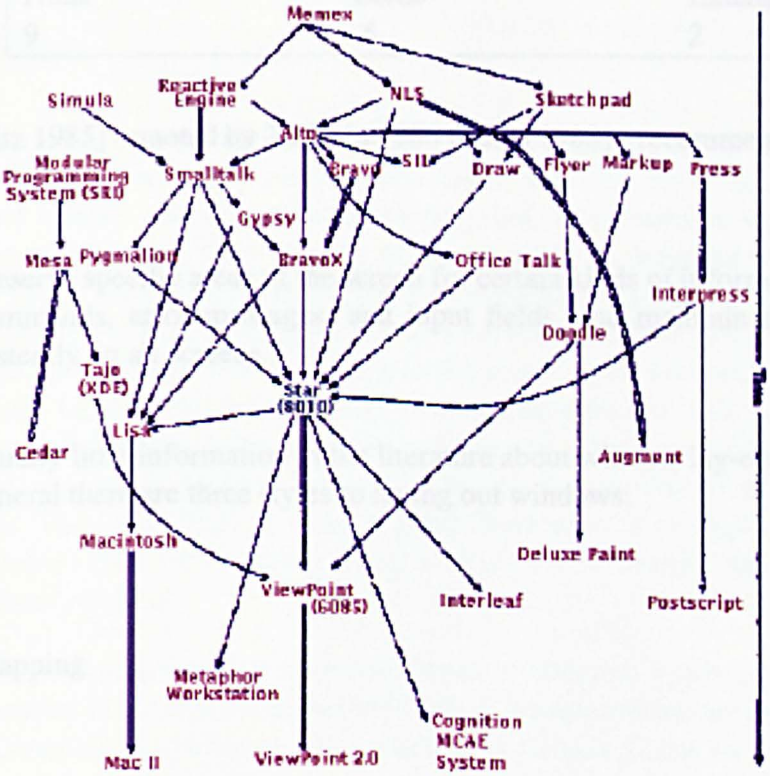


Figure A.1: The figure summarises the history of Star development until 1989.

The placement of windows and icons in many systems has frequently been haphazard and many screen design guidelines have not been supported by reference to the research that has been done. Tullis [Tullis 1988] provides a table of such guidelines where one of his classifications is "References to Empirical Data" and his results are summarised below:

Screen Design Guidelines: References to Empirical Data		
None	Some	Extensive
9	5	2

Galitz [Galitz 1985] - quoted by Tullis in 1988 [Tullis 1988] - recommends that designers should

> "...reserve specific areas of the screen for certain kinds of information, such as commands, error messages, and input fields, and maintain these areas consistently on all screens."

There is actually little information in the literature about window lay-out and icon placement. In general there are three styles to laying out windows:

1. Tiled
2. Overlapping
3. Mess

Furthermore, glancing through the literature it is clear from most screen dumps that number 3, the complete mess, is the most commonly chosen. Although all windowing systems provide complete control over window location, at start-up and by dragging during a session, it is perhaps surprising that more has not been done to establish conventions for screen lay-out. Overlapping windows in a stack is, of course, used to show the order of multiple windows which is particularly suitable for applications that use software such as Hypercard [AppleComputerInc. 1991], Supercard [AldusCorp. 1991] or Windows. Lynch [Lynch 1994a] claims that the current overlapping techniques are limited and relatively crude and superimposing objects is still the primary means of visual structuring in GUIs. Overall, and particularly when large screen, high resolution work stations are considered, little effort seems to have been placed on spatial location as a visual user interface cue. Similarly there is no generally accepted convention for icon location.

Many individual users, of course, design their own screen lay-out and use it consistently but this is not an acceptable strategy, although resorted to by many system designers, as there is no reason why users should know the best means of consistently laying out screens. Establishing a platform independent, internationally accepted convention may be a much cheaper option to aid users than using some of the more novel and expensive capabilities (e.g. virtual reality) just becoming available and affordable.

A.3.2 Motion

The ability to move high resolution, digitally generated images in real time is a relatively recent capability on commonly available machines, although the integration of high resolution video images with computer technology has been available rather longer. The disadvantage of video has been that the images need to be prepared in advance, which makes this sort of technology suitable for education and training systems and for some database, including hypermedia, applications where the database contents do not change frequently. Diaper [Diaper 1988] doubts that the extent to which users of the full virtual reality systems that include headsets, datagloves, treadmills, etc. will like to be "plugged in" to their computers; he suggests games and training environments are exceptions. Of course, what may be an esoteric application to one person may be mundane to others and Hubbard [Hubbard 1992] provides a good illustration of an application, molecular modelling, where full virtual reality using a visual stereo headset and six-dimensional mouse is almost essential.

Restricting consideration here of virtual reality to a VDU and a joystick or mouse then it remains necessary to question what one would wish to represent to users via moving digital images in a commercial environment (e.g. excluding games). In one way the use of moving images initiated by a computer system has disadvantages similar to those of speech user interfaces [Newell 1984, Diaper 1986, Diaper 1988] in that moving images are transitory and unscannable. If a major element of future user interfaces involving computer initiated moving images then users will have to attend to the screen so as not to miss events. It is also likely that such interfaces will be slow, not because of processor speed but because there are cognitive limits to the desirable speed of motion. The psychophysics of such interfaces will be complicated because the visual system is highly tuned to the detection of movement and there will be a complex interaction between resolution, spatial and temporal step sizes. Multiprocessor architectures such as UNIX will make the reliability of the time between successive images particularly difficult to control. Providing undo or repetition of image movement to overcome the problem of users missing events could lead to a graphical meta-dialogue, similar to those often advocated in natural language processing interfaces, this could be very slow.

The literature provides several technical papers and reports concerned with "program animation". Interestingly, and symptomatic of current available knowledge, many of the programs are rarely concerned with real motion.

Although virtual reality appears to offer new visual human-computer interface capabilities, one might suspect the use of full virtual reality systems in many mundane applications. Also, it remains necessary to question what one would wish to communicate to users via moving digital images in a commercial environment. Outside of specialised and often very narrow application domains, virtual reality seems to be another "solution looking for a problem". Keeping in view of the disadvantages initiated by moving images, it is reasonable to suggest that, at a minimum, what is needed is the establishment of a metaphor or some solutions for moving images user interfaces.

A.3.3 Depth

Over the centuries artists have developed more than a dozen systems of representing the third dimension of depth on a two dimensional surface. Given that many users in several environments do not wish to wear special glasses, let alone virtual reality head-sets, it is nevertheless still possible to represent depth on flat VDUs. High resolution graphics may not be essential for all depth representation methods, but it enhances all and is essential for many. It is also worth recalling that using binocular disparity (stereopsis) alone to indicate depth produces some odd effects. For example, objects in complex scenes often appear as projected on flat planes rather than the continuous depth experience normally enjoyed. Some people have complained of eye strain and the binocular disparity induced effect of depth is quite slow to appear; often the depth effect increases over some tens of seconds of viewing. The use of red-green filters is not recommended for the 8% male and 0.5% female population who are red-green colour blind [Brown and Cunningham 1989].

To date, depth has not been widely used as a visual cue to support users in performing tasks as it could be. Brown and Cunningham's [Brown and Cunningham 1989] proposal with respect to colour and depth "depth cueing" has been discussed in Chapter 2 (possibly related to the "atmosphere effect" [3])

The fundamental issue, of course, is what to use depth for. One option might be to use depth to portray some aspect of time, for example, as a user's session progressed objects on the scene that were unused could be represented as further away. In other words, this induced

[3] Depth cueing by varying the colour saturation of objects has been used for centuries by landscape artists to suggest great expanses of space within paintings. The painters noticed that the atmosphere causes faraway objects to lose colour saturation and visual details, and that mimicking these natural visual effects could produce paintings with the illusion of great distance [Lynch 1994a].

effect of depth might be useful in the user interface, where important foreground objects should be rendered in strongly contrasting hues. Background or unimportant elements may be visually shifted back away from the user by rendering them in desaturated colours or different colours with same saturation or shades of gray[4]. Furthermore, the psychological literature strongly suggests that time is a very important retrieval cue in human memory yet it has been rarely exploited: an exception is MEMOIRS [Lansdale *et al.* 1989]. Alternatively depth can be used to represent importance, frequency, or something else useful to others.

[4]Gray backgrounds are often gentler on the eye than the bright paper-white backgrounds very much seen in both the Windows and Macintosh interfaces [Lynch 1994a].

Appendix B

Colour Guidelines

B.1 A Sample Set

Literature available on colour as a powerful visual cue was reviewed as part of the work presented in this thesis. This review indicated that there exists several hundreds of guidelines and styleguides on using colour for different media applications, for example computer and print media. These guidelines and style guides were proposed based on human colour perception, colour reproduction in particular on computer displays, cultural and physiological emotions associated with colour and exploiting the use of colour as a powerful visual cue in coding information. Further, the recommendations have considered special needs of colour deficient people (8% of male and 0.4% female are colour deficient [MacDonald 1999]).

This section provides a sample of 50 guidelines on using colour in computer systems.

<u>**MacDonald's Guidelines (1999):**</u>

Colour Vision:

- The relative luminance of saturated colours follows the spectral luminous efficiency function, not the spectral hue order.
- Avoid adjacent areas of strong blue and strong red in a display to prevent unwanted depth effects (colours appearing to lie in different planes).
- Never use the blue channel alone for displaying fine details such as text or graphics. Do not use, for example, blue text on a black background or yellow text on a white background.
- Areas of strong colour and high contrast can produce afterimages when the viewer looks away from the screen, resulting in visual stress from prolonged viewing.

- Do not use hue alone to encode information in applications where serious consequences might ensue if a colour-deficient user were to make an incorrect selection.

Colour perception:

- Surrounding colours, field size, and viewing conditions can all change the appearance of colours.
- Where accurate visual judgement of a colour is necessary, the surround should be a neutral mid-grey to avoid unwanted perceptual colour changes.
- We can describe colour more meaningfully in terms of the perceptual dimensions of lightness, hue, and colourfulness than in terms of device signals.

Display technology:

- RGB display signals are device-dependent, and the colour they produce will generally differ from one display to another.
- When you need to render colours accurately, use a calibrated display and gamma correction software for best results.
- Some colours may be impossible to reproduce exactly if they lie outside the display colour gamut.
- Use perceptual colour models based on CIE uniform colour spaces instead of the simplistic colour models based on device-dependent RGB signals.

Design Principles:

- Treat colour design as part of a user-centred design process.
- Get it right in black and white, then add colour sparingly.
- Use colour for association and differentiation of a design's elements.
- Choose a harmonious palette of colours for use throughout an application.
- Unify each design by using common thematic colour(s).

Graphic user interfaces:

- Use colour consistently throughout all screens in an application.
- Use strong colours in small details only, such as icons and graphical indicators.
- Use a limited palette of colours and offer predefined harmonious combinations.

Advertising:

- Use bright, highly saturated colours to grab attention, but not for prolonged viewing.
- Take advantage of psychological associations of colours.
- For Web graphics, use colours from the browser-safe palette.

Text:

- Ensure good legibility by providing adequate contrast between text and background.
- Avoid coloured text on a coloured background.
- Use the metaphor of a highlighter pen to draw attention to areas of text.

Information:

- Use colour in conjunction with other visual variables for effective presentations.
- Use strong colours sparingly on or between muted background tones.
- Limit the number of colours in nominal coding to seven or fewer.
- Use natural or application-related associations for ordinal coding.
- Always use a colour key or a scale with a colour-coded display.
- Use colour transparency to show overlays of related structures.

Visualisation:

- Don't use colour that doesn't support or add to the meaning of the information displayed.
- Use colour in monitoring applications to indicate changes of state.
- Use colour saturation to depict depth layering and priority of object categories.
- In modelling applications, use only enough colour to create a realistic effect.

Imaging:

- Always use neutral grey background when displaying colour images.
- Put a narrow white border around an image stabilise its colour appearance.
- Use colour management software when accurate colour reproduction is required.

Summary - five golden rules

- Take account of human visual needs and expectations.
- Conform to the colour conventions for the application.
- Design the screen layout considering all available visual variables.
- Be consistent in the use of colour throughout all screens in an application.
- Use colour sparingly, never more than is necessary for the task.

Guidelines taken from Cornell University Web site:

http://ergo.human.cornell.edu/AHTutorials/interface.html

- use attention grabbing techniques cautiously (e.g. avoid overusing 'blinks' on web pages, flashing messages, 'you have mail', bold colours etc.)
- don't use more than 4 different font sizes per screen use serif or sans serif fonts appropriately as the visual task situation demands
- use colours appropriately and make use of expectations (e.g. don't have an OK button coloured red! use green for OK, yellow for 'caution, and red for 'danger' or 'stop')

- don't use more than 4 different colours on a screen
- don't use blue for text (hard to read), blue is a good background colour
- don't put red text on a blue background
- use high contrast colour combinations
- use colours consistently
- use only 2 levels of intensity on a single screen
- Use underlining, bold, inverse video or other markers sparingly on text screens don't use more than 3 fonts on a single screen

Guidelines from Shneiderman (1983, 1997) -

- Don't overuse! Use between 3 - 7 colours at the most in one application. four at the most on any single display. Colour coding should support the task (for example, used to flag important status changes)
- Be consistent within the application and with user expectations (e.g., popular stereotypes in the West are that red = emergency, green = OK)
- Be careful about colour combinations (need good contrast yet not too bright, don't mix saturated blues and reds, don't put yellow text on a white background, etc.).
- Also be aware that some members of the population may be colour blind – try to design for monochrome.

Guidelines taken from http:csweb.cs.bgsu.edumaner

- Use colour only as an enhancing, secondary form of information
- Use a limited set of colours
- Use system metrics for all display elements (such as colour settings and fonts)

Appendix C

Experiment 1 - Pilot Study

C.1 Results

Data recorded from the first eight users was analysed in the first part of Experiment 1. As this part was designed as a pilot study, only the time taken to complete the test tasks was analysed (see Figure C.1). Other statistics, such as error counts, were not subjected to analysis at this stage. A Mann-Whitney U test was used to determine the difference between the time scores of control and experimental groups. The test results did not show a significant difference between the times (Mann-Whitney U value of 5 and a one-tailed probability of 0.243).

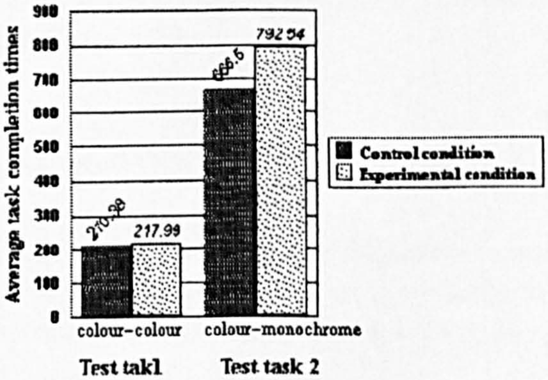


Figure C.1: Test conditions and average task completion times in pilot study (Experiment 1).

C.2 Conclusions

The difference (see Figure C.1) between the Test Task 2 completion times between the two conditions was not significant. This could be due to the small sample size. This pilot study was conducted mainly to determine the sensitivity of Test measures. The analysis of the results establish the sensitivity and it was decided to complete the full experiment with additional test users.

Appendix D

Experiment 1: Tasks and Tools

D.1 Specific Action/Object Selector

This tool takes one or more prose descriptions of the task as input and provides a suitable environment for the user to construct and manipulate lists of objects and/or actions, from the given description, through the different command buttons.

Input - Prose description of the task

Output - Lists of objects and/or actions

The following subsections describe the inputs and outputs for both the training and testing tasks in Experiment 1 (see Chapter 6).

D.1.1 Prose Description of the Task - Training

Living things are best thought of as being in either the animal or plant kingdom. There are five main groups of vertebrates. Of particular relevance is whether they are endothermic(warm-blooded) or ectothermic(cold-blooded). Mammals include gerbils, rabbits and dolphins.

D.1.2 Resulting list of objects - Training

The tool allows users to construct an object list of animals and birds :

Object List		
gerbils	rabbits	dolphins

D.1.3 Prose Description of the Task - Testing

Living things are best thought of as being in either the animal or plant kingdom. All the species discussed here are animals whether mammals, birds, reptiles, amphibians, fish or invertebrates. There are five main groups of vertebrates. Mammals include gerbils, rabbits, bats, dolphins and deer. Most of the cage and aviary birds are pet birds and they include lovebirds and cackatoos. Other groups of birds are wild birds and birds of prey. Falconiformes are the examples of birds of prey and include vultures, falcons and hawks. Some examples of falcons and hawks are peregrine, saker and goshawk, sparrow-hawk respectively. Reptiles come under vertebrates and include tortoises and snakes. Amphibians are classified into adults with tails for example newts, adults without tails for example frogs and some tropical legless species. Invertebrates are classified into joint-limbed animals, insects, centipedes and millipedes, crabs and crayfish. Insects include various species of butterfly, cockroaches and grasshoppers. Most of the species mentioned above are exotic pets with a few exceptions.

From the above description, the test users were asked to construct a list of birds and animals following the sequence of operations mentioned below:

- Step 1
  The initial screen includes the following four command buttons and a selection window containing the prose description of the task together with associated line index numbers.

  - LOAD : To load a file of text

  - QUIT : To exit the selection session

  - SAVE : To save the displayed text

  - EDIT : To create lists of specific objects and/or objects from the task description

- Step 2
  Select any line by clicking the left mouse button once on one of the lines. Selection is confirmed by the black striped line.

- Step 3
  Clicking on the EDIT button displays a selection dialogue box labelled 'Select Action/Object'. The box contains the currently selected line and two list environments, one for handling specific actions and the other for specific objects.

- Step 4
  To construct an object list, select[1] the object part of the selected line in the selection dialogue box, for example gerbils, and then click the SPECIFIC OBJECT button. Note that the selected object appears in the list with a numeric identifier, and square brackets delimit the object in the selection line. The numeric identifier indicates the correspondence between objects and actions. To deselect any object, select the object from the list and click the DESELECT button. The action list is constructed in the same way. **NOTE: Current task includes construction of the object list only.**

- Step 5
  Click the DONE button to confirm the selection of objects and/or actions from the chosen line. To delete the newly selected actions and/or objects click the CANCEL button.

- Step 6
  Clicking the DONE button brings the participant back to the initial screen , where he can select another line.

- Step 7
  In a similar way the lists of selected objects and/or actions are constructed from the entire set of lines.

- Step 8
  Selections are reflected in the set of lines and current text selections can be saved by clicking on the SAVE button. Confirmation of the operation appears over the SAVE button.

- Step 9
  Click on the QUIT button to end the session.

NOTE: If QUIT is clicked prior to SAVE , a dialogue box will appear. Click again the SAVE option to ensure no work is lost.

1

- **To select a word:** Move the pointer to the specific word and click the left mouse button twice in rapid succession
- **To select multiple words:** Move the pointer to the beginning of part of the line and drag the mouse towards the end

D.1.4 The Outcome of the Test Task

Figure D.1 shows the resulting object list.

Object List		
gerbils	rabbits	bats
dolphins	deer	lovebirds
cockatoos	peregrine	saker
goshawk	sparrow-hawk	tortoises
snakes	newts	frogs
butterfly	cockroaches	grasshoppers

Table D.1: Object list constructed using SASO in Experiment 1.

D.2 Task Descriptive Hierarchy Editor

The Task Descriptive Hierarchy Editor (TDH) provides powerful tools for both the construction and manipulation of hierarchy like structures that describe a task in terms of objects, actions and the relationships between them. The first step to build a task descriptive hierarchy is to construct a very high-level description of the task that describes all of the object and/or action combinations. Once the high-level specification has been determined, the rest of the hierarchy can be constructed. See Figure D.1.

Input - Lists of objects and/or actions

Output - Hierarchy of objects and/or actions.

Example :

Input Object list for the Training Task

Object List		
gerbils	rabbits	dolphins

Resulting Hierarchy in the Training Task

```
Living things
|___ plants
|___ animals
     |___ vertebrates
          |___endothermic(warm-blooded)
          |     |
          |     |___ mammals
          |             (3 dolphins,)
          |             (2 rabbits,)
          |             (1 gerbils,)
          |
          |___ectothermic(cold-blooded)
```

D.2.1 TDH Operations

The task involves constructing a hierarchy of living things.

- Initially, the task descriptive hierarchy editor along with two windows showing specific action and/or object lists will appear on the screen.
- At first, the editor has only one node, labelled 'ROOT-LABEL'. It is from this node that all subsequent node labels are attached. For this reason the mnemonic label for the root node should be determined and assigned , for example it can be 'Living things'.
- All lower-level nodes should be attached to this node using the ADD CHILD or INSERT PARENT operations.

- Initially, the very high-level description of the task is constructed using the ADD CHILD and RENAME operations.
- If any mistakes are made, they can be corrected by selecting the node (object/action) to be deleted and then clicking on the DELETE button.

The sequence of operations to build the high-level hierarchy are described in Tables D.2, D.3 and D.4.

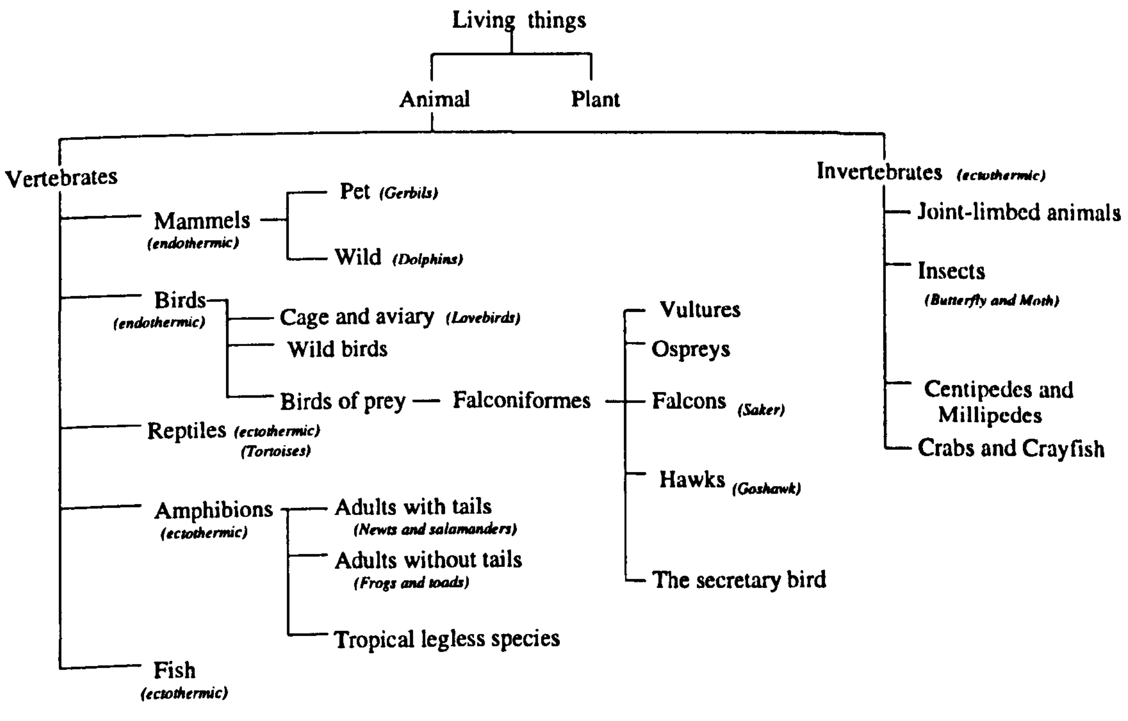


Figure D.1: Hierarchy of Living Things

Successful use of the above mentioned tool results in the required task descriptive hierarchy. Save the hierarchy by clicking on the SAVE button. Click on DUMP to dump the screen to a file and on QUIT to quit the TDH tool.

Input Object List for Test Task 2

gerbils	rabbits	bats	dolphins	deer	lovebirds
cockatoos	peregrine	saker	goshawk	sparrow-hawk	tortoises
snakes	newts	frogs	butterfly	cockroaches	grasshoppers

Sequence of operations	*Command buttons*	*Result*
Change ROOT-LABEL to Living things	1. Select ROOT-LABEL 2. Click on RENAME	A dialogue box appears
	3. Move the cursor into the text input area of the dialogue box and Type in 'Living things' 4. Click on 'DONE'	ROOT-LABEL is changed to Living things
Add child node 'animals' to 'Living things'	5. Select 'Living things' 6. Click on ADD CHILD	A node BLANK-LABEL appears
	7. Select BLANK-LABEL 8. Follow 2-4 and rename BLANK-LABEL to 'animals'	Node 'animals' appears instead BLANK-LABEL
Add child node 'plants' to 'Living things'	Follow 5-8	'Living things' contains two child nodes: 'animals' and 'plants'
Add child nodes 'vertebrates' and 'invertebrates' to 'animals'	Follow 5-8	'animals' with 'vertebrates' and 'invertebrates' as child nodes appear
Add child nodes 'mammels','birds', 'reptiles','amphibians', and 'fish' to 'vertebrates'	Follow 5-8	'vertebrates' node appears with 'mammels','birds' 'reptiles','amphibians', and 'fish'
Add child nodes 'cage and aviary', 'wild birds' and 'birds of prey' to 'birds'	Follow 5-8	'birds' node appears with 'cage and aviary', 'wild' and 'birds of prey'
Add child nodes 'adults with tails', 'adults without tails' and 'tropical legless species' to 'amphibions'	Follow 5-8	'amphibions' node appears with 'adults with tails', 'adults without tails' and 'tropical legless species'
Add child node 'falconiformes' to 'birds of prey'	Follow 5-8	'birds of prey' node appears with 'falconiformes'
Add child nodes 'vultures','falcons' and 'hawks' bird' to 'falconiformes'	Follow 5-8	'falconiformes' node appears with 'vultures' 'hawks' and 'falcons'

Table D.2: The sequence of operation in Test Task 2 (Experiment 1)

Sequence of operations	*Command buttons*	*Result*
Add child nodes 'joint-limbed animals', 'centipedes and millipedes' 'insects','crabs and crayfish' to 'invertebrates'	Follow 5-8	'invertebrates' node appears with 'joint-limbed animals','centipedes and millipedes','insects' 'crabs and crayfish'
*To delete any node	1.Select the node 2.Click on DELETE	A dialogue box appears to confirm DELETE
	3.Click on YES of dialogue box	The node and it's children are deleted
Add object 'gerbils' to 'pet'	1.Select 'pet' 2.Click on ATTACH SO	A selection box containing all objects appears
	3.Select 'gerbils' from object list 4.Click on ATTACH 5.Repeat 1-4 to attach other objects to 'pet'	'pet' is attached with 'gerbils'
Follow 1-5 to add all other objects to respective nodes		
To add an action to any node	6.Choose the action 7.Click on ATTACH SA	A selection box containing all actions will appear
	8.Select the action from action list 9.Click on ATTACH	node is attached with the action
Follow 6-9 to add all other actions to respective nodes		
*To delete any object or action	10.Select node from which objects or actions are to be detatched 10.Click on DETATCH SA or DETATCH SO	Selection box containing all currently attached objects or actions appears
	11.Select object or action 12.Click on DETATCH	Objects or actions are detatched

Table D.3: Continuation of Table D.2.

Resulting Hierarchy from Test Task 2

The order of adding child nodes to the hierarchy is shown below:

```
                                                         * order *
Living things
|___ plants.............................................. FIRST
|___ animals............................................. FIRST
    |___ vertebrates..................................... SECOND
    |    |___ mammals.................................... THIRD
    |    |      (11 dolphins,)
    |    |      (8 gerbils,)
    |    |___ birds...................................... THIRD
    |    |    |___ cage and aviary....................... FOURTH
    |    |    |      (13 lovebirds,)
    |    |    |___ wild birds............................ FOURTH
    |    |    |___ birds of prey......................... FOURTH
    |    |         |___ vultures......................... FIFTH
    |    |         |___ falcons.......................... FIFTH
    |    |         |      (18 peregrine,)
    |    |         |___ hawks............................ FIFTH
    |    |                (19 goshawk,)
    |    |
    |    |___ reptiles................................... THIRD
    |    |      (21 tortoises,)
    |    |___ amphibians................................. THIRD
    |    |    |___ adults with tails..................... FOURTH
    |    |    |      (23 newts)
    |    |    |___ adults without tails.................. FOURTH
    |    |    |      (24 frogs)
    |    |    |___ tropical legless species.............. FOURTH
    |    |
```

```
|     |___ fish...................................... THIRD
|
|___ invertebrates.................................. SECOND
      |___ joint-limbed animals...................... THIRD
      |___ insects................................... THIRD
      |      (28 butterfly)
      |___ centipedes and millipedes................. THIRD
      |___ crabs and crayfish........................ THIRD
```

D.3 Testing for Colour Blindness

Ishihara's [Ishihara 1998] test uses different colour plates in determining the colour deficiency of people. Figure D.2 shows one of the plates [URL 1998b]. Interpretation of the plate by the person determines the colour vision of the user. The user with normal vision can read the figure for the number 5. The colour-blind can hardly read the plate.

The usability studies reported in this thesis have made use of this plate as part of an approach employed in determining the colour deficiency of a test user. The approach was based on knowledge elicitation concepts (see Chapter 5).

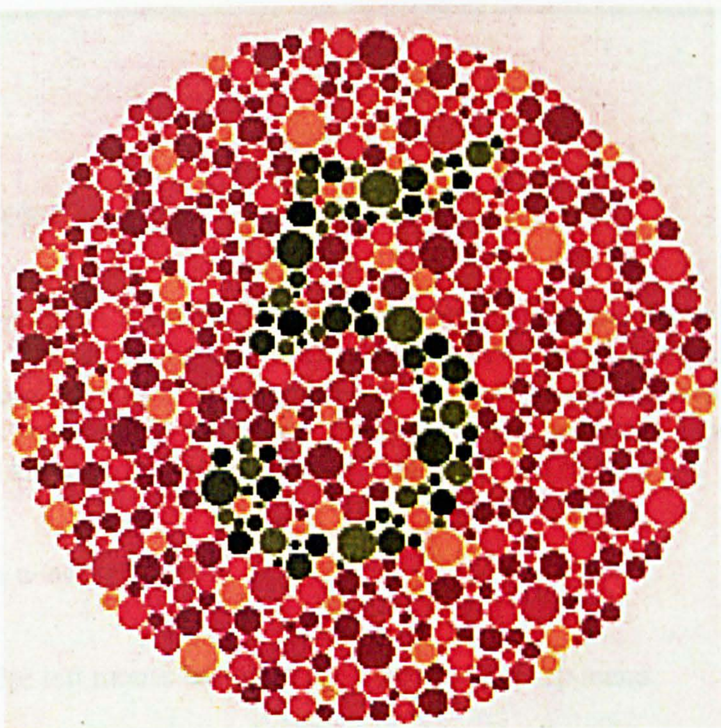


Figure D.2: Testing for colour blindness - people with normal vision read the figure for the number 5, but the colour-blind can hardly read it.

Appendix E

Experiment 2: Tasks and Tools

E.1 Xaero

Task1 of Experiment 2 includes training and testing participants on the use of 'Xaero' application. This is a program to edit and animate a virtual world using objects, for example sphere, cylinder, cuboid, plane etc. The task involved animating a simple world using objects for example sphere and cylinder. The sequence of operations to build the virtual world are described in below (see Tables E.1 and E.2).

- Open a window 'Xterm'
  - Use left mouse button to open the root pop-up menu.
  - Drag the cursor to 'Window ops' icon.
  - Drag the cursor to 'New Xterm' and release mouse button.
    Result: Xterm window appears.
- Type **program name (for example, Test Task 1)** in the text window
  Result: The common front-end window with different buttons appears
- Fill in your NAME and EXPERIMENTAL CONDITION (for example, CONDITION = MONOCHROME) text fields.
- To Use 'XAero' :

Note:

- On wrong selection either click on respective 'delete buttons' or repeat the sequence of operations.
- Test Users are advised to ignore the details of other command buttons at this stage.

Sequence of operations	*Command buttons*	*Result*
Select Xaero	Click on 'Simulation - aero' button	A screen labelled AERO appears
Select the object Sphere	1. Click on 'insert new object' 2. Select sphere from the pull down menu 3. Click on 'alu' from the material list 4. Move 'G' slider to downwards (little bit) until the dark colour 5. Click on 'apply color' 6. Change radius to 5.00 7. Click on 'fixed'	Object window appears Sphere appears in xyz planes Dark colour in colour box Colour of sphere changes to dark colour
Select cylinder	8. Click on 'select next object' 9. Repeat 1-2 with cylinder 10. Click on 'iron' 11. Click on 'apply color' 12. Change dimensions to 2 and 2 13. Click on 'fixed'	Cylinder appears in xyz planes
Group the objects	14. Click on 'group objects' 15. Click on 'apply' and then 'done'	A small dialogue box appears at top right corner of the screen

Table E.1: The sequence of operations in Test Task 1 (Experiment 2).

<u>**To close the session**</u>

- Use left mouse button to open root menu
- Drag the mouse to 'Exit X-Windows' and release the button

Sequence of operations	*Command buttons*	*Result*
Link objects using Spring	16. Click on 'link objects' and select 'spring' from pull down menu 17. Change spring length to 2.1 18. Click on 'apply' 19. Click on first button in the top row 20. Click on 'apply' 21. Click on 'next one' and 'apply' 22. Repeat 22-23	A dialogue box labelled 'Spring Parameter' appears A box for 'position link point' appears A box for 'next object to link' appears A box for 'position link point' appears
Add forces to objects	23. Click on 'Add force' and select 'wrt world' from pull down menu 24. Repeat 22-23 25. Set force to 200 and click on 'apply'	A box for 'force point' appears A box for 'setting force' appears
See Animated world	26. Click on 'Animation' 27. Click on 'Play' button 28. Click on 'Stop' button 29. Click on 'Close'	Animation window appears Results animated world Animation stops Close animation
Save animation	30. Click on 'input/output' and select 'save world' 31. Type a file name and click on 'ok'	A dialogue box appears
Exit AERO	32. Click on 'AERO' and click on 'Quit' 33. Click on 'Quit'	A dialogue box appears Quit confirmed

Table E.2: Continuation of Table E.1

E.1.1 Testing - Test Task1

1. Open a new text window and type **program name** at the prompt.

2. Use Xaero to animate a simple world with

- **Objects :** sphere (material - aluminium; select dark colour; radius = 5.0), cuboid (material - gold; dark colour; dimensions (2.4, 2.5, 1.2)) and cylinder (material - iron; dimensions (2.2, 2.0))
- **Spring length 2.1**
- **Position and force points** - along x (right direction, clockwise)
- **Force - 200N**

E.2 Xalarm

Xalarm is an interactive alarm clock for the X window system. Test Task 2 of Experiment 2 uses Xalarm and the associated task involves specifying the time and date for the alarm. The test users were instructed to follow the sequence of operations mentioned below (see Tables E.3 and E.4).

- Open a window 'Xterm'
  - Use left mouse button to open the root pop-up menu.
  - Drag the cursor to 'Window ops' icon.
  - Drag the cursor to 'New Xterm' and release mouse button.
  - **Result: Xterm window appears.**
- Type **the program name** in the text window
  Result: The common front-end window with different buttons appears.

- Fill in your NAME and EXPERIMENTAL CONDITION (for example, CONDITION = MONOCHROME) text fields.

- To Use 'XAlarm' :

Sequence of operations	*Command buttons*	*Result*
Set the time	1. Click on 'Set Alarm' 2. Type in the time e.g. 5:15pm	A dialog box appears
Set the date	3. Click on 'Date' 4. Type in the date e.g. +1	+x indicates the alarm time for x number of days.
Set warnings	5. Click on 'Warnings' 6. Enter '5' text input box	
Confirm time	7. Click on the left most button three times till 'Confirm time' appears. 8. Click on 'Save' 9. Click on 'Confirm time'	A dialog box that includes 'confirm time' and 'Alarm Call' message appears Saving is confirmed The dialogue box disappears
On warning	10. Click on 'dismiss'	Warning window disappears
At the alarm time : read and quit	11. Click on 'Quit'	The alarm box at the specified time and date disappears

Table E.3: The sequence of operations in Test Task 2 (Experiment 2).

E.3 Xwais

Xwais is an interface to one of the mostly used tools for exploring the resources of the internet namely WAIS (Wide-Area Information Servers) system and allows the user to select and modify WAIS questions and sources. The task to be performed includes a simple search of sources of information following the sequence of operations mentioned below.

NOTE: Participants are requested to ignore the details of other command buttons at this stage.

<u>To close the session</u>

Sequence of operations	*Command buttons*	*Result*
Select Xwais	1. Click on Questions & Sources	XWAIS window that includes questions and sources appears
Open question 'bitnet'	2. Click on 'bitnet' of question list and on 'Open' button	A dialog box appears
		Available documents are listed
Add the first six documents to the list * On wrong selection **	3. Click on the first document 4. Click on 'Add Document' button 5. Repeat 3-4 to add the remaining five documents * Select the document added to the list * Click on 'delete document'	 Selected item appears in the document list
Save the document	6. Click on 'Save...' button 7. Type a file name in text input box 8. Click on 'Save' button	A dialogue box appears Documents are saved
Quit the question session	9. Click on 'Quit' button	
Read and save the contents of a source	10. Click on any source in the list 11. Click on 'open' button of source list 12. Click on 'Save' button	A dialogue box that contains the source information appears Contents are saved and the box disappears
Quit Xwais	13. Click on Quit twice	End of Xwais session

Table E.4: Continuation of Table E.3

- Use left mouse button to open root menu
- Drag the mouse to 'Exit X-Windows' and release the button

E.3.1 Testing - Test Task2

Open a new text window and type **the program name.**

- **Use XALRM to set the alarm at current time+8 minutes for the next three days**
- **Use XWAIS with the question 'Internet'**

Appendix F

Experiments 1 and 2 in Retrospect

F.1 Analysis

The important question is: why did Experiment 2 (see Chapter 7) fail to produce a colour consistency effect? Further, why didn't the experiment replicate at least the colour-monochrome effect found in Experiment 1? The analysis did not provide any satisfactory reasons.

Task	*Experiment*	*Types of Operations Application used*	*Average Task completion times*	*Colour Effect?*
Task 1.1	1	Search and Identification (SAO:object selection)	236	No (confirms with predictions)
Task 1.2	1	Search and Identification List construction (TDH:hierarchy editor)	632	Yes (confirms with predictions)
Task 2.1	2	Object Manipulation (Xaero:animation tool)	404	Yes (against predictions)
Task 2.2.1 Task 2.2.2	2 2	Object Manipulation (Xalarm:updating time) Search and Identification (Xwais:document search)	181	No (confirms with predictions)

Table F.1: Types of operations and approximate task completion times in Experiments 1 and 2

The failure prompted the analysis of the test tasks used in both Experiments 1 and 2. In general, tasks included different types of operations such as search and identification, constructing lists and/or hierarchy and object manipulation. An attempt is made to classify the test tasks in terms of these three types of operations (see Table F.1). Primarily, Tasks 2.1, and 2.2.1 required the users to manipulate visual objects (for example, the parameters of a sphere) through spatially distributed interface widgets such as pushbuttons. Tasks 1.1, 1.2 and 2.2.2 required the users to search and identify target items. Further, Task 1.2 required the users to associate target objects with target classes. To this end, Tasks 1.1, 1.2 and 2.2.2 could be considered cognitively more complex relative to Tasks 2.1 and 2.2.1. This analysis may lead to the suggestion that the types of operations executed by the user would affect his/her performance.

The times in the fourth column of Table F.1 represent the minimum average completion times in all the test conditions for each task. For example, the minimum value for Task 1.1 is computed by comparing the average completion times in both control and experimental conditions of Experiment 1 (See Figures 6.5 and C.1). Tasks 1.1 and 2.2 took the shortest time followed by Task 2.1 and then Task 1.2, which took the longest. The results consistently confirmed the predictions in cases of small task size. They were however inconsistent with the predictions in cases of longer task sizes. That is, Task 1.2 produced a significant colour effect as per the predictions and Task 2.1 failed to produce the same. These observations may allow one to infer that task size (measured in terms of task completion times) can probably influence the resulted effects against the predictions. Further analysis along these lines may lead one to explore the possibility that the effects (positive, no or negative) can be better detected if the tasks are of uniform size.

One could also take a different analysis. Three out of four tasks have produced results in accordance with the predictions (Tasks 1.1, 1.2 and 2.2). Task 2.1 is the exception in that it did not produce a significant colour effect. In the case of an inconsistent effect, the scheme failed to produce a significant one however it came out about 10% worse than the other (Figure 7.3). The result, if replicated over a large sample, probably would have been statistically significant.

Appendix G

WWW: Test Pages for Experiment 4

G.1 Experiment 4 - Pilot Study

G.1.1 Results

Data recorded from the first eight users was subjected to analysis. As this was a pilot study, only the time taken to complete the test tasks was analysed. Other statistics, such as non-optimal response counts, were not subjected to analysis at this stage. A Mann-Whitney U test was used to determine the difference between the time scores of explicit and implicit test groups. The average task completion times of the groups are shown in Figure G.1.

G.1.2 Conclusions

The difference (see Figure G.1) between the test task completion times between the two conditions is significant (Mann-Whitney U value of 3 and a one-tailed probability of 0.029). The results therefore establish the sensitivity of the test measures under the two test conditions. The objective of conducting the pilot study has been met. The results formed a basis for further testing of the impact of the Web colour scheme on user performance (see Chapters 9 to 12).

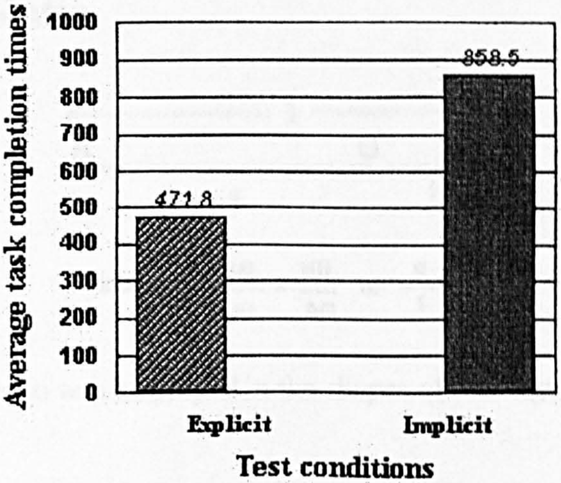


Figure G.1: Experimental groups and average task completion times in the pilot study of Experiment 4.

Appendix H

WWW: Test Pages for Experiment 5

H.1 Golden Ratio

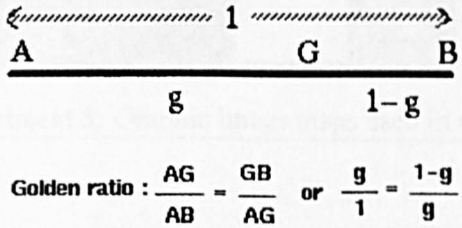


Figure H.1: Golden ratio was employed in the shapes of rectangular imags maps used in Experiment 5.

Throughout history, the golden ratio (ratio of length to width of rectangles of 1.61803 and more; see Figure H.1) has been considered as one of the most pleasing of visual features. Continuing research on the golden ratio has concluded that the ratio appears in natural phenomena as well as proportions that human eye perceives as beautiful [Amir 1996]. This golden ratio, also called 'divine proportion' or 'golden mean' or 'golden section', has been widely used in works of art and architecture as the classic ideal of beauty, and also in music. Some of the modern day designers have showed greater interest in employing the golden ratio in graphic works. This includes graphic forms designed for online use, such as the Web graphics [URL 2000].

Many books on oil painting and water colour [Elkins 1998, Allrich 1996] recommend that it is better to position objects not exactly in the centre of the picture but to one side or "about one-third" of the way across, and to use lines which divide the picture into thirds. This seems to make the picture design more pleasing to the eye and relies again on the golden section paradigm. The research employed this ratio in query pages 1 and 5, which consist of large rectangular image maps (see Figure H.2). The main purpose of using this ratio was to deliver aesthetically pleasing Web graphics.

H.2 Example Web Page Graphics in Experiment 5

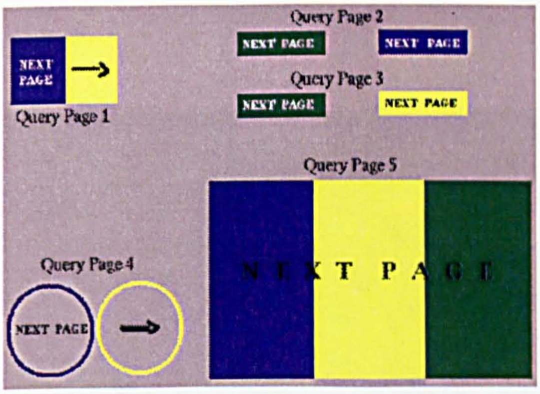


Figure H.2: Experiment 5: Graphic imags maps used in the test task.

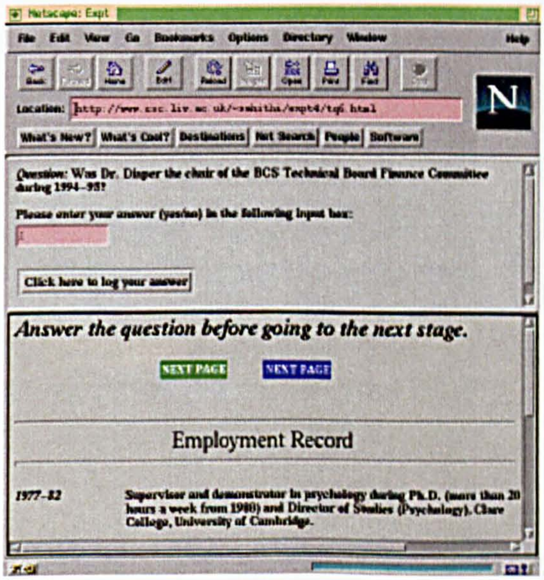


Figure H.3: Experiment 5: A scaled down screen dump of an example query page used in the test task.

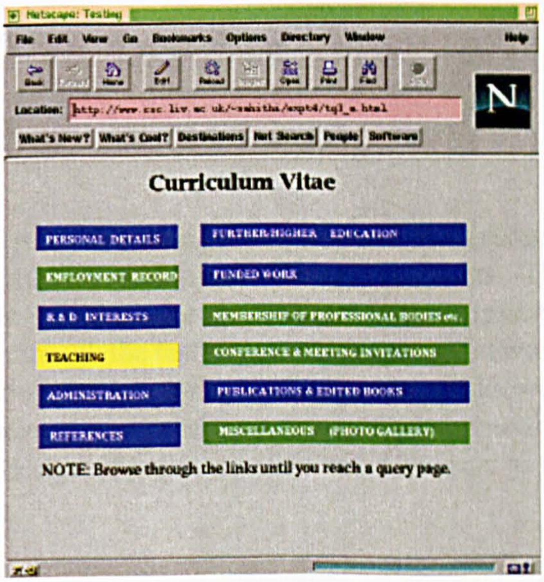


Figure H.4: Experiment 5: A scaled down screen dump of an example non-query page used in the test task.

Appendix I

Experiment 6: Pre-testing Web Graphics

I.1 Foreword

The pre-experimental study reported here attempts to design, analyse and modify the trial graphic designs for Experiment 6 based on users' reactions. The goal was to propose a set of trial designs that would encourage the users in the implicit condition to infer colour rules while browsing through the test pages. The experiment was paper based. The image dumps of candidate designs consisting of image map links were printed on the paper. They were tested on several users and their selection preferences were recorded. Based on their preferences the design of pages was fine tuned in order to help the implicit learning process.

I.2 Trial Graphics Design

Initially, the study involved trying out different graphic forms which can be used to represent hyperlinks in test pages. As in Experiments 4 and 5 (see Chapters 9 and 10), image maps in different shapes and sizes were used along with graphic buttons (Java applets). With an enormous number of possible trial designs, nine primary designs were selected as described in the following subsections. Design and selection of graphic forms was driven by the implicit learning of rules, i.e. it is expected that carefully formulated designs would produce better performance on every test trial in the implicit condition. Possible learning within the implicit condition was optimised during the early trials. The nine primary designs were implemented using the three colour scheme. The trial designs increased in difficulty from page to page. In particular, Trials 7 and 8 included antagonistic features which might to over-ride the learned three colour rules.

I.2.1 Nine Trials

I.2.1.1 Trail 1

Figure I.1: Trial Design 1

The three button design, with blue (B), green (G) and yellow (Y) as background colours, was the default graphic design (see Figure I.1). There were 6 possible sequences of ordering on the screen from left to right: (Y G B), (Y B G), (G B Y), (G Y B), (B G Y) and (B Y G). The geometrical order of placing the links on the screen should be selected from one of these. Trial 1 was intended to maximise learning of colour rules which requires an appropriate ordering of colours. Of the three colours, only B results in a long delay followed by an error message. Selection of G also results in a delay but the user is taken to the next trial. Y is the best selection, resulting in an immediate transfer to the next trial. The users selecting B first need to select another link to proceed to the next trial thus resulting in two selections. It was therefore expected that the users selecting the B as the first link would learn of its non-selectability. This necessitates a design in which users will prefer B link as the first link to select. Considering the left-right bias usually present among users, the design option BGY was thus considered best for optimising learning in Trial 1. There are four different possible sequences of selection, BG, BY, G, Y of which first two start with the B selection. 50% of possible user selections thus require two selections. The prediction was later confirmed in the pre-experimental phase in which a majority of subjects exhibited a left-right selection bias.

Repeating Trial 1 as Trial 9 allowed the learning within the experiment to be tested. It should be noted that the BGY order was not fully repeated within Trials 2-8. Trials 1 and 9 were identical with respect to size, shape and labels of the buttons. This allowed the users to make selections based on colour rules learned explicitly or implicitly.

Table I.1 lists all possible user responses in Trails 1 to 3. The string BG, for example, indicates user selection of a blue link first, followed by a green link. A single letter (Y or G) indicates successful selection and transfer to the next trial.

Trial 1 has four possibile responses and Trails 2 and 3 have two each. There were 16 possible patterns of response from a user (4*2*2) for the three trials. All except two patterns allow the selection of B in one of the three trials. The two patterns where B was completely absent was G,Y,G and Y,Y,G. If a person, for whatever reason, refuses to ever select B, then he/she can be assumed to effectively possess the B colour rule,

Trial 1	BG	BY	G	Y	[Left-Right Bias]
Trial 2	Y	BY			[Top-Bottom Bias]
Trial 3	G	BG			[Left-Right Bias]

Table I.1: Possible users' responses during the Trials 1 to 3

i.e. non-selectability of blue links. In the remaining 14 patterns, the presence of B was expected to produce a learning effect across the first 3 trials, if it was selected.

I.2.1.2 Trial 2

Figure I.2: Trial Design 2

Trial 2 (see Figure I.2) included two buttons displayed vertically, the top one with a yellow background and the bottom with a blue one. The trial used the vertical display in order to introduce users to the concept of display variability. Buttons had the same labels and were identical in shape and size. The Yellow button was placed on the top . The rationale was that given top-bottom selection bias, the top button was more likely to be selected. In the pre-experimental phase during paper based testing, the users exhibited a top-bottom bias. The Y link, placed on top, was the implicit condition users' expected selection. The Yellow colour rule, selectability of link and transfer to the next trial without delay, was thus introduced to users who had a top bias.

This bias could have been quite weak if the other possible alternative strategy, selecting the final item in the top-to-bottom list, in this case B, was chosen by the users. The pre-experimental data confirmed that only a minority of users followed this alternative strategy. Y and B were used because they were more extreme than G and B and the speed effect of Y selection was needed to be introduced to those who selected BG or G in Trial 1. The B selection gave an explicit error rather than just a delay.

I.2.1.3 Trial 3

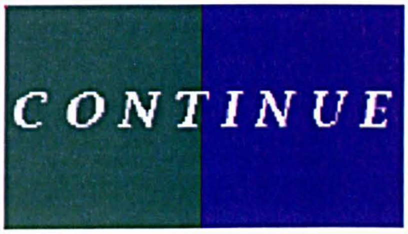


Figure I.3: Trial Design 3

The overall graphic design was regarded as an example of poor design because the layout did not indicate the buttons very clearly. However, similar designs were found on several Web sites.

This trial (see Figure I.3) used a horizontal bar; the background colour of the left half of the bar was painted in green and the right in blue. G was located in the primary left position. The labels on the left and right halves provided no information about the selectable parts of the bar. The use of two options again allowed greater experimental control of learning by users in the implicit condition.

There were several reasons for choosing the (G,B) pair for this trial. First, it was necessary for implicit condition rule induction.

Second, if B was the first in Trial 1, Trail 3's order would reinforce perceived randomness of Y,G,B link order. Third, the explicit condition users who had a simple Y rule (i.e. always choose Y) were forced to apply a non-Y rule. It should be noted that the implicit condition users who responded with BY or Y in Trials 1 or 2 may also have inferred the Y rule.

I.2.1.4 Trial 4

The graphic image (see Figure I.4) differed from the image in Trial 3 in size and background colours. The first part of the image was painted with a blue background and the second part, which was larger than the first, with a gray background and yellow border.

Trial 4 was explicitly intended to allow implicit condition users to discover the border rule: that borders have the same effect as background colour. Second, the trial was an attempt to make users biased towards incorrect B responses by making it appear more visually

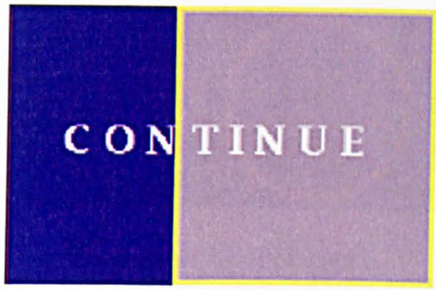


Figure I.4: Trial Design 4

powerful than the Y border; placing B on the left also contributed to this bias through the left-right bias effect. As such, this was a strong test of explicit learning.

I.2.1.5 Trial 5

Figure I.5: Trial Design 5

This trial (see Figure I.5) included three buttons of the same sizes and labels, placed horizontally (left to right) next to each other. The first and the last buttons used the same background colour B, where as the middle one used G.

The presence of two B options on the same page, introduced to users to the notion that the same colour may be used more than once (see Trial 8). The usage of the same layout as in Trials 1 and 9 prevented a response in Trial 9 based on only one previous instance of the layout.

B and G were chosen so as to prevent the Y rule dominating the experiment. It should be noted that forcing a non-Y response required only 2 colour options; Trial 5 allowed this with three locations. This would make users not associate number of colour options with number of location options. this trial allowed the measuring of user's learning in Trial 3 i.e. G vs. B.

I.2.1.6 Trial 6

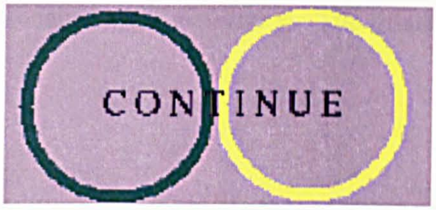


Figure I.6: Trial Design 6

Trial 6 (see Figure I.6) tested the border rule in a novel graphic design. This included two circles identical in size and with a background colour not yellow, green or blue. The left circle used a green border and the right one used yellow.

G and Y were chosen to test the Y rule, since both G and Y were selectable links. Implicit condition users might by now have had adequate experience to infer the colour rules. For example their possible selection pattern (ignoring B selections) in the first 5 trials could be Y, Y, G, Y, G. In the worst case (i.e. users who preferred G to Y) would have had the pattern: G, Y, G, Y, G. Users under the implicit condition, who followed Y-Y-G-Y-G or G-Y-G-Y-G might have learned that both G and Y had an identical effect, without realising that Y was faster than G. With this rationale, G and Y were chosen to allow users in the implicit condition have the to learn the differences in speed associated with G and Y. Trial 6 therefore provided a strong test of colour learning. The ordering of G and Y was based on the left-right bias effect as established in the pre-experimental phase.

I.2.1.7 Trial 7

Figure I.7: Trial Design 7

The graphic design comprised a long white bar with a green border and a small bar with a blue background colour, located to the right of the white bar.

Trial 7 tested the B rule against the border and G rules and thus complemented the design in Trial 6. As with Trial 4, the strong visual effect of B was designed to over-ride the B rule and make users choose the non-selectable blue link. The tick-mark was intended as a strong antagonistic cue to cause error generation. The graphic design was a variant of Trials 3 and 4.

I.2.1.8 Trial 8

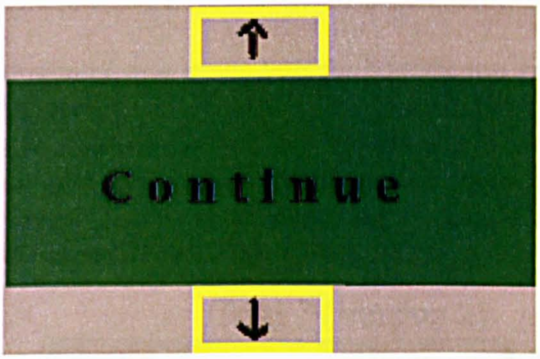


Figure I.8: Trial Design 8

In Trial 8 (see Figure I.8), the top and bottom squares were drawn in gray with an yellow border. The big rectangle in the middle was painted with a green background.

Trial 8 re-tested the Y and border rules as in Trial 6 and attempted to over-ride them with the red arrow on the middle rectangle. This made the G link most likely to be selected (this user choice was confirmed by the pre-experimental data).

One design option considered was to make the Y parts indicated only by a Y border, this would make the test even harder. This option was not selected. Trial 8 was still a very strong test of explicit group learning; the implicit group were likely to choose the G part unless they had inferred the Y rule, so it was a very strong test for both the conditions.

I.2.1.9 Trial 9

Same as Trial 1.

Table I.2 summarises the graphics design of all nine trials. The first row shows the number of colour links in each trial. Except for trials 1 and 9 (which were identical) all had only two colour options. The first and last had all three colours. Which colours are present in which trials is also indicated. For example, Trials 3, 5 and 7 consist of 2 colours: Blue and Green. The second row illustrates colour rules. The Y links have the highest preference followed by G and B links. The third row specifies the location of target colours Y and G. The last row specifies the graphics layout in each trial. Except Trials 2 and 8, all had a horizontal layout.

I.2.2 Trials 7 and 8 - Multiple Versions

Trials 7 and 8 were considered as a strong tests for both the test groups, in particular the explicit group, and therefore it was decided to design and analyse multiple versions of these two trials before deciding on the final set. The aim was to choose the most complex versions of Trials 7 and 8 and then examine the effect of the three colour scheme on users' task performance. Consequently, the initial study used three different versions of Trials 7 and 8 each as shown in Figures I.9 and I.10.

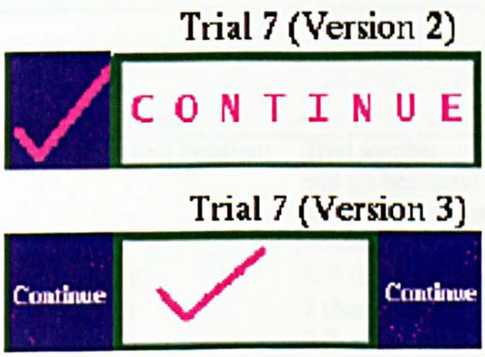


Figure I.9: Multiple versions of Trial 7 in the Pre-experimental study (Experiment 6)

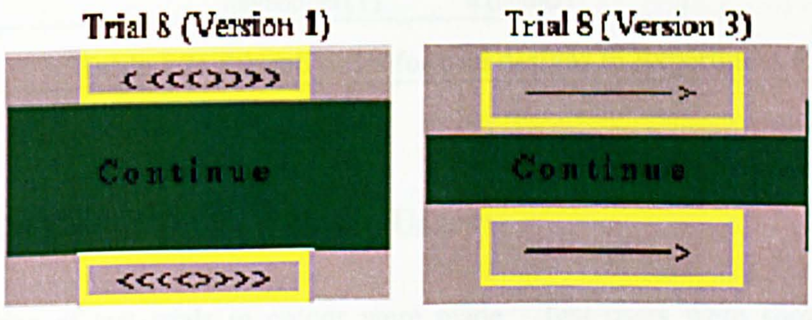


Figure I.10: Multiple versions of Trial 8 in the Pre-experimental study (Experiment 6)

Having designed the trial graphics, the study proceeded to design and implement the Web pages using the graphics as image maps which represented hyperlinks. The results of testing these versions are described in Section I.4.

Query page number	*Number of links*	*Colour choices*
2,4	Two	B or Y
3,5,7	Two	B or G
6,8	Two	G or Y
1,9	Three	B or G or Y
	Colour rule in each trial	
Degree/order of preference: Y(High) G(Medium) B(Low)		
Across pages: *The location of the target link (either yellow or green) was changed randomly.* *Trials 1, 9 and 3, 7 maintained Y-G-B order.*		
Graphics layout	Target link location:	Trial number and (in brackets) colour feature used
Horizontal	right (Y)	1, 9 (background)
Vertical	top (Y)	2 (background)
Horizontal	left (G)	3 (background)
Horizontal	right (Y)	4 (background)
Horizontal	middle (G)	5 (background)
Horizontal	right (Y)	6 (border)
Horizontal	left (G)	7 (border)
Vertical	top/bottom (Y)	8 (border)

Table I.2: Stimuli order for trial designs in Experiment 6.

I.2.3 Testing of Trials with the Users

Hard copies of test trials in colour were made. Test users were shown each screen and asked to point out where they would place the cursor first to make a selection. In cases of trials with three colour options, test users were asked for their second choice. Users' preferences were recorded and tabulated on separate data sheets for later analysis. Each user was interviewed to try and ascertain why the user had selected a particular colour/location. Their opinions about the graphic design were also noted. The study did not intend to explain to the users anything about the colour design prior to testing. As mentioned in previous sections, the reason for conducting these interviews was to modify the graphic designs for the purpose of achieving good task performances in the final testing phase.

Overall, the study devoted about one hour to each user. Since the study was not measuring users performance, test users were not required to have a high degree of Web proficiency. The only requirement was to recruit people who were familiar with the basic concepts of Web browsing, such as identifying and selecting hyperlinks.

I.2.4 Test Users

Thirteen users, aged between 20 - 27 years, participated in this study. All were Liverpool university students and none had prior knowledge of the use of colour in user interface design. They had not participated earlier either in paper based experiments or in experiments testing colour design. All had normal or corrected to normal vision. The Web experience of the users varied from four months to two years.

I.3 Data Analysis - Modifying Trial Designs

User selection criteria and opinions were analysed for the purpose of modifying trial designs where necessary. The data allowed some determination of the actual psycho-logic, for example left-right (as observed in Trials 1, 3, 6 and 9) and top-bottom (as observed in Trials 2 and 8) biases and users' preferences for certain colour locations. It should be noted that this user modelling as logic was for the purpose of experimental design, not analysis.

The principal interest, however, was to determine users' preferences in selecting the differently coloured links. The results are summarised in Table I.3. The table lists all the options of selection and number of users (in brackets) choosing each option. For trials with three colours (1 and 9) two colours of choice and the preferred order are shown. For the others, only one colour is selected. For example, in Trial 1 five users selected BG, i.e. blue link was the first choice and green link was the second choice. For three users it was other way round (option GB). Only two users selected Y as the first choice and then green as the second (option YG). In Trial 4, a majority of seven users selected Y as the first choice. Five users selected B as the first choice and none selected G.

I.4 Trial Designs for Experiment 6

Selection criteria, such as order of selecting links, adopted by a majority of users was almost identical to the criteria assumed while designing the trial graphics (see Section I.2.1). In other words, the study demonstrated that the recommended set of trial designs was consistent with the expected of users' biases, primarily left-right and top-bottom, and thus would allow implicit learning of colour rules as planned. Among different versions of Trials 7 and 8, Versions 7.1 and 8.2 were considered the most appropriate designs for testing the explicit colour learning because a majority of users selected the wrong target regions in the image maps. In summary, it was decided to use trial designs 1 to 6, 7.1, 8.2 and 9 as they were without further modifications later.

Trial design	*Selection order - and number of users*	*Analysis*
1	BG (5) - BY (3) - GB (3) - YG (2)	Majority preferred BG i.e. more chances to learn B rule
2	Y (9) - B (4)	More chances to learn Y rule
3	G (9) - B (3) - Nil (1)	More chances to learn G rule
4	Y (7) - B (5) - Nil (1)	More chances to learn Y rule
5	GB (8) - BG (3) - Nil (1) GB: G then B located right BG: B then G located left	More chances to learn G rule
6	G (9) - Y (3) - Nil (1)	More chances to learn G rule
7 (Version 1)	Tick mark (10) - G (3)	Most appropriate design to test users' colour learning
7 (Version 2)	Tick mark (8) - G (4) - Nil (1)	
7 (Version 3)	G (8) - Right tick mark (3) then G (2)	
8 (Version 1)	G then Y (8) - Y (5)	
8 (Version 2)	G then Y (10) - Y (3) i.e. Y (bottom) - Y (top)	Most appropriate design to test users' colour learning
8 (Version 3)	G then Y (6) - Y (5) - Nil (2)	
9	BG (6) - BY (4) - GB (2) YG (1)	More chances to learn B rule

Table I.3: Analysis of users' preferences in selecting differently coloured links: Experiment 6 (Pre-experimental study).

Appendix J

The Web Style Guide

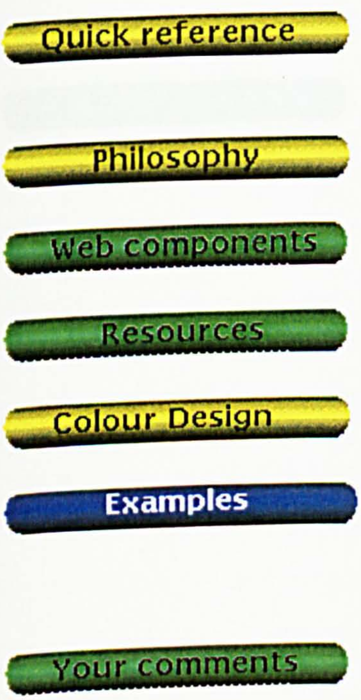
J.1 Introduction

The style guide for the proposed Web colour product is available on the Web at http://www.csc.liv.ac.uk/ ~sahithi/ styleguide.html

The style guide comprises seven sections as shown in Figure 12.1, each summarising relevant sections from Chapter 4. This appendix presents hard copies of some sections only. Details of all the sections can be found on the Web.

J.1.1 Home Page

A Style Guide for Background/Border Colour



This Style Guide describes one experimental use of colour for aiding users browsing Web pages. The functional design is very simple. Links, either as text or graphics, have a coloured background or border that informs users how accessible is the information at the end of the link (Blue - Green - Yellow: least to most accessible). Foreground colour is not used, although it is necessary to ensure that foreground and background colours are compatible.

You must have at least: Netscape 3.1, or a similar Web browser.

The Style Guide can be accessed via the topic button on the left. While we have tried to provide browser safe colours and have tested the example pages with a number of browsers (e.g. Netscape 3.1 onwards, Internet Explorer, etc.) and on both UNIX workstations and PCs, we would be interested in feedback where our colours do not work; of course, provide your system details as far as possible. The default text link colours are changed for the purpose of research.

We are interested in whether you find the functional design useful and whether it offends any aesthetic you may have. We welcome your comments on the contents of the Style Guide itself.

Philosophy

Colour functionality is one of the important topics, closely related to User Interface (UI) Design and implementation problems, that has been widely neglected to date. Colour is widely used in Web Page Design (WPD). To date, colour has been studied by Web designers more from its technical usage, for example: how to produce different colour pallets, what are the safe colour options for different browsers etc. Apart from the technical research, colour usage has mainly been limited to crude grouping and little effort is devoted to exploiting functional aspects of colour.

This style guide mainly focuses on the use of colour as an important display feature in WPD. A plausible colour design was proposed and discussed with respect to some common problems reported by Web users.

What this Style guide is not

The overall goal of this Style manual is to propose a plausible colour scheme that demonstrates the possibility of developing more functional Web pages using colour. It is not to provide definite solutions to all of the issues associated with both functionality and colour in WPD.

The advice here is aimed at the practical concerns of bending and adapting colour to real purposes it was never intended to serve.

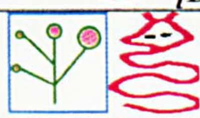
- Style guide: Home
- Quick reference
- Philosophy
- Resources
- Colour Design
- Examples
- Your comments

Web Page Components - Identifying Web Links

In general, a Web page includes a number of different components; for example, text, hyperlinks and graphics. While text and graphics convey some information, hyperlinks are mainly used for Web traversal. The chief power of Web browsing comes from hyperlinks, which are simply pointers in one page to elsewhere in the same page or to another page entirely.

Links are the selectable components in a Web page and they are generally distinguished from the non-link components by a piece of underlined text, graphics such as buttons, bullets, images (animated, static, dynamic) and icons, and colours or reverse video (black on white, for instance).

For example, a link is shown:

by a piece of underlined text	Text Link ***[Do not try this link]***
as a button	,  ***[Do not try these links]***
as a bullet	 ***[Do not try this link]***
as an image	 ***[Do not try these links]***
as an animated image or icon	 ***[Do not try this link]***

A recent survey of Web pages indicated that 50 out of every 100 links are graphical. A major problem with graphic links is that users often find it difficult to identify the graphics that are selectable from non-selectable ones. Consider the following example. The Web page includes six buttons out of which only four are selectable. All the buttons look alike and there is no immediate visual support for the user in identifying selectable ones at a glance.

See Figure 4.1

As the organisation of information developes in Web sites with very dense text, graphics and links, selection becomes more difficult for users. The problem is further intensified with the use of imagemaps.

Image Maps

An image may be an important navigational aid or locator, or may be the most effective means to organise page layout, or help set readers' expectations about a site. In today's Web browser implementations, image maps, the selectable graphics used on Web pages, present a rather poor user interface. Unless the image itself has well-delineated "active" regions, there is no clear indication of where a user should click.

Consider the following image as an example. No doubt it takes some time to load but do you easily identify the link part which takes you to another page? How much time does it take to find out the selectable area in this image map?

See Figure 4.2

Yes, it takes a considerable amount of time to find out the selectable and non-selectable regions. Pages with such image maps fail dismally in "hooking" a reader because of the frustration incurred when people are subjected to spending more time in navigating and visualising Web content, and are given tantalising but unwanted glimpses of content they need. The ambiguity about where to click or what to select on an image map will further magnify the problem. In order to avoid frustrations;

Users should be informed about the selectable regions, in general, selectable components as quickly as possible. How can they be informed?

Colour can be used as an important physical cue to clearly delineate selectable components and regions in a Web page. *A plausible* **colour design** *is proposed to serve as a navigational aid and support users in identifying such Web components, at a glance, before making any request.*

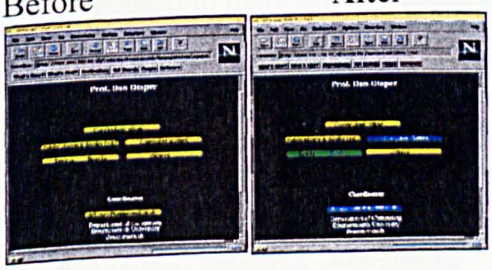
Style guide: Home | Quick reference | Philosophy
Web components | Resources | Colour Design

Your comments

Colour Design Implementation - Examples

Example-1:

Before After



Example-2:

Before After

Bournemouth University's School of Design, Engineering & Computing (DEC) Web pages.

J.2 Application of the Three Colour Design in Real Web Pages

Courses | Staff | School of DEC | Academic Departments | Research

DEPARTMENT OF APPLIED PSYCHOLOGY

HEAD OF DEPARTMENT: Gerry Griffin

The relationships between people and technology are complex and varied, and impact on all areas of life, from the home to sophisticated economic-generation activity. Understanding of technology in supporting human activities is essential to our development, especially those aspects concerned with information management.

The Department's main focus is on the BSc (Hons) Applied Psychology and Computing course and the range of research activity undertaken reflects this commitment. Members of the Department are interested in many aspects of the relationship between individuals and technology and the topics currently being pursued include:

- Group processes
- Human and machine speech production
- Artificial intelligence
- Multimedia communication
- Natural language processing
- Occupational psychology
- Vision

Students are provided with the knowledge and skills necessary for finding fulfilling employment and to that end we concentrate on maintaining a balance between the academic and the practical. The Department is well provided with modern hardware and software in purpose-built laboratories.

Xanadu - Experimental Web Server

Courses | Staff | School of DEC | Academic Departments | Research

This page is maintained by Daniel Vine, and was last updated 25 June1997.

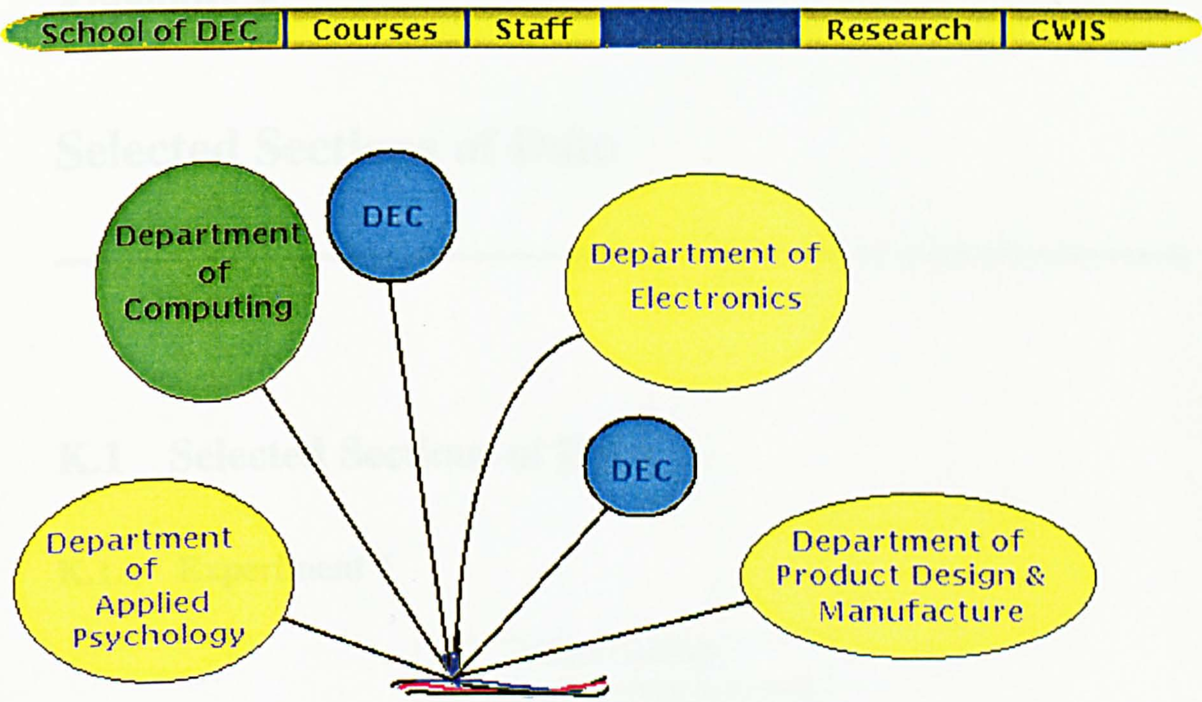
DESIGN ENGINEERING and COMPUTING

Staff | Research & Consultancy

Academic Departments | Courses

A Brief by the Head of School

This page is maintained by Daniel Vine and was last modified 4 February 1998.



Send comments to Daniel Vine (dvine@bournemouth.ac.uk)

Last updated 5 February 1997.

Appendix K

Selected Sections of Data

K.1 Selected Sections of Data

K.1.1 Experiment 1

Control Condition)		
Data (Completion times in seconds)		
Subject	Task1	Task2
1	193.72	700.28
2	203.12	520.69
3	232.40	589.72
4	212.29	637.22
5	255.35	747.16
6	237.90	509.43
7	183.51	657.77
8	286.09	731.63
9	287.45	563.09
10	284.83	662.12

Experimental Condition		
Data (Completion times in seconds)		
Subject	*Taskl*	*Task2*
1	210.72	591.85
2	200.65	685.11
3	250.01	800.33
4	210.56	630.49
5	203.98	671.79
6	243.05	754.48
7	320.72	1030.22
8	212.21	713.67
9	270.32	1122.51
10	234.40	841.74

K.1.2 Experiment 2

Group0: Explicit *Data (Completion times in seconds)*		
Subject	Task1 *Consistent Colour*	Task2 *Consistent Colour*
1	526.9	240.3
2	512.9	288.4
3	337.6	195.0
4	472.1	163.2
5	449.5	176.5
6	232.4	112.5

Group1: Implicit) *Data (Completion times in seconds)*		
Subject	Task1 *Consistent Colour*	Task2 *Consistent Colour*
1	498.4	245.3
2	274.6	175.7
3	421.3	199.5
4	471.6	228.6
5	397.5	216.2
6	393.7	238.4

Group2) *Data (Completion times in seconds)*		
Subject	Task1 *Inconsistent Colour*	Task2 *Consistent Colour*
1	460.1	152.7
2	412.3	215.1
3	512.5	204.1
4	539.5	224.7
5	390.3	209.8
6	514.1	206.4

Group3) *Data (Completion times in seconds)*		
Subject	Task1 *Monochrome*	Task2 *Monochrome*
1	372.2	169.7
2	463.6	219.4
3	478.9	214.3
4	366.5	213.6
5	421.4	186.2
6	321.8	145.0

K.1.3 Experiment 5

Implicit Condition (With Delays)) Data (Completion times in seconds)		
Subject	(Questions 1,2,3,4,5,6)	Total
I1	96,99,126,125,171,178	795
I2	59,110,48,91,152,128	588
I3	83,146,48,36,77,102	492
I4	32,79,45,113,148,63	480
I5	62,30,70,64,82,101	409
I6	130,135,57,68,136,57	583

Explicit Condition (With Delays)) Data (Completion times in seconds)		
Subject	(Questions 1,2,3,4,5,6)	Total
E1	13,27,28,19,61,41	189
E2	19,25,34,14,47,52	191
E3	20,31,29,7,51,28	166
E4	9,30,31,10,39,31	150
E5	17,22,38,31,52,20	180
E6	8,28,19,21,49,38	163

Implicit Condition (Without Delays)) Data (Completion times in seconds)		
Subject	(Questions 1,2,3,4,5,6)	Total
I1	51,34,106,80,106,133	510
I2	14,45,28,46,87,83	303
I3	58,81,48,36,77,57	357
I4	32,59,45,68,83,63	350
I5	17,10,50,19,82,101	279
I6	86,70,37,23,71,57	344

Explicit Condition (Without Delays)) Data (Completion times in seconds)		
Subject	(Questions 1,2,3,4,5,6)	Total
E1	13,7,28,19,16,41	124
E2	19,5,34,14,47,30	149
E3	20,11,29,7,31,28	126
E4	9,10,31,10,39,31	130
E5	17,2,38,31,52,20	160
E6	8,8,19,21,49,18	123

K.1.4 Experiment 6

Implicit Condition (With Delays)) *Data (Completion times in seconds)*		
Subject	(Questions 1,2,3,4,5,6,7,8,9)	Total
I1	93,89,106,123,111,133,96,78,102	931
I2	61,92,41,81,102,118,72,69,84	84
I3	103,146,78,71,98,107,119,147,139	1008
I4	32,79,45,113,148,63,97,103,59	739
I5	47,30,70,64,82,86,109,73,45	606
I6	106,75,57,67,99,51,88,61,49	653
I7	92,84,28,59,149,102,126,102,67	809
I8	110,137,89,63,59,151,121,96,79	905

Explicit Condition (With Delays)) *Data (Completion times in seconds)*		
Subject	(Questions 1,2,3,4,5,6,7,8,9)	Total
E1	34,27,45,56,68,71,88,62,32	483
E2	42,47,53,59,52,83,125,94,35	590
E3	20,31,29,17,69,78,107,41,27	419
E4	35,36,31,32,48,100,113,88,23	506
E5	39,32,67,48,47,83,97,56,24	493
E6	28,31,41,43,53,121,91,81,22	511
E7	23,12,29,16,60,94,69,52,16	371
E8	14,38,30,23,54,61,66,21,9	316

Implicit Condition (Without Delays)) *Data (Completion times in seconds)*		
Subject	(Questions 1,2,3,4,5,6,7,8,9)	Total
I1	48,44,41,78,46,113,31,78,37	516
I2	41,47,21,36,37,98,7,49,39	375
I3	58,101,58,71,78,87,54,127,74	708
I4	32,79,25,68,83,63,77,83,39	549
I5	27,30,15,19,62,66,44,53,25	341
I6	86,30,37,22,34,51,68,61,49	438
I7	92,39,8,59,84,82,61,82,67	574
I8	90,92,44,63,30,131,56,96,79	681

Explicit Condition (Without Delays)) *Data (Completion times in seconds)*		
Subject	(Questions 1,2,3,4,5,6,7,8,9)	Total
E1	34,27,25,56,48,71,68,62,32	423
E2	42,47,33,59,32,63,105,94,35	510
E3	20,31,9,17,4,78,87,21,27	294
E4	35,36,11,32,28,100,93,68,23	426
E5	39,32,47,48,27,63,77,56,24	413
E6	28,31,21,43,23,121,71,81,22	441
E7	3,12,9,16,40,94,49,52,16	291
E8	14,38,10,23,34,61,46,21,9	256

Questions 7 & 8 - User First Selection *Y-Yellow;G-Green;B-Blue* *Top, Bottom, Tick indicate image locations*			
Implicit:Q7	Explit:Q7	Implicit:Q8	Explicit:Q8
I1(Y-Top)	E1(Y-Top)	I1(Tick)	E1(G)
I2(G)	E2(Y-Top)	I2(Tick)	E2(G)
I3(G)	E3(G)	I3(Tick)	E3(G)
I4(G)	E4(G)	I4(G)	E4(G)
I5(G)	E5(Y-Top)	I5(Tick)	E5(G)
I6(Y-Top)	E6(Y-Top)	I6(G)	E6(G)
I7(G)	E7(Y-Bottom)	I7(Tick)	E7(G)
I8(Y-Top)	E8(Y-Top)	I8(Tick)	E8(G)

Glossary and Abbreviations

Achromatic	Without the quality of Colour [Robertson 1979] .
Ambient light	This is the light surrounding your display. It may be normal window light, sunlight, incandescent light, fluorescent light, and so on. The ambient light makes a difference in the way colours appear on the screen.
AOL	America OnLine.
Browser-safe colours	The 256 colours that do not shift between platforms, operating systems, or most Web browsers [Weinman and Heavin 1997a].
CRT	Cathode Ray Tube. An electronic vacuum tube that is used to display pictures containing alphanumerics, text, graphics or images in VDUs [Robertson 1979].
CUI	Colour User Interface.
DM	Direct Manipulation.
FTP	File Transfer Protocol. A method of transferring files over the Internet.
GIF	Acronym for Graphics Interchange Format. It is a bit-mapped graphics file format. GIF is commonly used in designing images for the Web because it uses an efficient compression method.
GUI	Graphical User Interface.
HCI	Human-Computer Interaction.
HLS model	Hue, Lightness and Saturation model.

HSB model	Hue, Saturation and Brightness model.
HSV model	Hue, Saturation and Value model.
HTML	Hypertext Markup Language, a text encoding language (written in SGML) used in creating Web pages.
Hue	The chromatic component of the colour, the quantity that distinguishes between colours such as red, green and blue. **See Saturation and Brightness.**
Hyperlinks	Pointers to different types of documents in a Web page. Documents contain information in the form of text, graphics, video or audio. **See also Resource.**
Hypermedia	Like hypertext, except that the modules can contain animation, sound, video, and other media as well as text and graphics. Hypermedia could be considered as multimedia modules connected with hypertext links.
Hypertext	Modules of text and simple graphics stored electronically and accessed by electronic links between modules. For online documentation you usually must supplement the hypertext links with other methods of access.
Internal consistency	addresses the consistency of the computer system structure itself
Internet	A large world-wide interconnected network of computers which allows exchange of information using several protocols, such as the World Wide Web. **See WWW.**
Image map	The selectable graphics used in designing Web pages. They represent one form of hyperlinks.
ISDN	Integrated Services digital Network.
JANET	Acronym for Joint Academic Network representing UK's academic and research network.
JSD	Jackson Structured Diagram.
Lightness/Brightness	How dark or light a colour is.

Lisa	Apple introduced Lisa in 1983, a product similar to the STAR in terms of its user interface. **See also STAR.**
Luminance	This component of the picture information is responsible for detail, shapes, and shadings.
OA	Office Automation system.
RGB	In general, colours are specified by their Red, Green and Blue components.
RGB model	A dominant colour model for computer applications. This model specifies colour as an additive combination of Red, Green and Blue primaries.
Monochrome	A display of a single colour.
Multimedia	Mixture of text, graphics, animation, video, sound, music, and perhaps other media, usually in a linear sequence.
Non-optimal Response	In this thesis, an undesirable action performed by a test user in an empirical study was counted as a non-optimal response.
Resource	A resource is a document containing information in various forms. Resourcse vary in content and size, and are downloaded whenever a user activates/ selects hyperlinks that comprise locations of URLs with a keystroke or a mouse click. **See also hyperlink and URL.**
Saturation	The purity of a colour, which ranges from neutral gray to pure colour. It can be described as a quality of colour related to the range of wavelengths in the light being observed, or to the degree of white in the colour. For example, CRT red is saturated, white is completely unsaturated; pink is medium saturated [Robertson 1979, Macintyre 1991].
SGML	Standard Generalised Markup Language, specified by ISO. A descriptive language which allows textural material to be encoded in a manner that specifies the logical structure of a document with a separate description of how this document should appear on the page. HTML is an SGML encoding scheme.

Softcopy Electronically provided information in a book format. Softcopy may be searched and read from the computer but more often is searched on the computer, printed out, and read from paper.

STAR The Xerox 8010 office information system is known as STAR. It was probably the first comprehensive DM system intended for a business application in an office environment [Smith *et al.* 1983]. **See DM.**

Test User A participant in an empirical study.

Tritanope A person who is blue colour blind.

UID User Interface Design.

URL Universal Resource Locator.

VDU Video Display Unit.

VIBGYOR A colour spectrum generally associated with a rainbow: Violet-Indigo-Blue-Green-Yellow-Orange-Red.

White Point The image on a display is a combination of red, green, and blue signals. All lights, including display light, have a white point, which is the measure of the colour content of the light. White point is measured in degrees Kelvin, and it sets the foundation for the other colours on a display. If the white point is high, colours have a bluish tinge. If the white point is low, colours have a slightly reddish tinge.

WIMP Windows, Icons, Menus and Pointing Devices.

WWW Acronym for World Wide Web. A set of protocols for the exchange of hypermedia information over computer networks. It provides a simplified user interface (both graphical and text) to other Internet tools. The physical location and format of a document have no bearing on the methods used to access the Web.

WPD Web Page Design.

WYSWYG What You See is What You Get.