ADAPTIVE INTERFACES: COGNITIVE STYLES AND PERSONALITY CHARACTERISTICS AS DETERMINANTS OF SUPPORT

Mary-Ann Forrest

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

The University of Leeds,
Department of Psychology
September 1988

Abstract

The main purpose of the thesis was to examine the human computer interface from a user-centred viewpoint in order that it adapt to the characteristics of the user in terms of his/her personal style (cognitive and learning styles, personality traits). The relationship between personal style and performance, and comprehension of the system's operations are explored in order to highlight the characteristics that are most likely to predict the type of user that is the most successful learner. The manner in which the interface can support the user was explored in terms of the type of environment best suited to each individual, as was the use of colour (non-qualitative) to provide feedback as to the user's location within a system. Qualitative colour was also examined in relation to performance and the potential of using qualitative colour coding to enhance the user's response.

Two methodologies were employed: applied (Studies 1-3) and naturalistic (Studies 4 and 5). The former attempts to simulate an environment which forces the subjects to use information processing strategies as if they were in a naturalistic environment.

The findings of Studies 1 and 2 indicated that the Embedded Figures Test (EFT) is a good predictor of success in a simulated word processing environment. Field independent subjects were found to consistently perform better than their field dependent counterparts. In Study 1 it was

expected that colour would act as a support to the user and would therefore benefit field dependent users. The results suggested that colour hindered performance but this may be due to overuse. Study 2 also investigated the usefulness of coding information. Colour and shape coding were found to be equally good as cues for remembering over no form of coding.

Many of the studies which have investigated the use of colour to enhance performance have relied on redundant colour coding of arbitrary symbols. Study 3 explored the usefulness of qualitative colour coding and meaningful stimuli to determine the effects on performance and the potential of using qualitative colour coding at the interface. The results showed that there are no differences in response time for qualitative colour coding, non-qualitative colour coding and no colour coding. The lack of significant findings suggested that position of the stimuli may have confounded the results.

In a naturalistic study (Study 4) it was possible to tentatively predict from the subject's personal style, the type of software environment most suited to him/her through observation of his/her behaviour in addition to the types of problems encountered by subjects and their frequency of occurrence.

Study 5 attempted to relate natural language and comprehension to personal style. The findings suggested that question-asking techniques do not appear to be a contributor to successful learning. Additionally, the most efficient learner in terms of the number of tasks that s/he was able to complete and scores on a comprehension test, was the redundant holist and holist. Serialists were found to be slower than holists, and thus the design of the study may have contributed to these findings.

Taken together, the studies suggest the type of user who tends to be most successful in interacting with computers in terms of behavioural data such as number of keystrokes, frequency of asking questions, frequency of asking for help, time spent reading the manual etc, in terms of the misunderstandings experienced and in terms of their comprehension of system operations. The use of colour as a support to less able users was not borne out in the study and may be due to the care that is required in employing colour at the interface. Colour as an associative cue may be effective in comparison to no colour cue. However, other forms of cueing may be equally effective, such as shape coding.

Acknowledgements

I would like to thank my supervisor, Dr N.P. Sheehy for his help and support during my research programme. Thanks also to Professor A.J. Chapman for his comments and criticisms of the thesis.

Many thanks to members of the Psychology Department, from Technical and Research staff to Postgraduate and Undergraduate students, for their participation in the experiments. Special thanks to Susan Beardsell, Simon Gilburt and Gillian McKay for their "patience" over three traumatic weeks of learning to use the word processors.

CONTENTS

Chapter 1 Conceptual Overview

1.1 Introduction

This chapter defines some important concepts concerning the human computer interface. It also attempts to integrate different aspects of the interface within a framework. The main thrust of the thesis is to demonstrate the desirability and feasibility of implementing an adaptive interface taking the "user" as the starting point, and using information about the user to modify the interface dynamically.

At this point it is necessary to describe what is meant by 'the interface' and what it means to modify it. (I shall leave a description of an adaptive interface until later). The interface is taken to mean everything through which the user communicates with the system. It includes the mode of communication - be it command language or graphical. In addition, the type of support facilities available to the user together with all aspects of the screen (such as the use of colour), keyboard and any pointing devices are included.

It is also necessary to consider the different meanings of the term "user". Users can be distinguished in terms of their level of expertise. For example, the three categories normally referred to are novice, familiar and expert. In addition, we need to be aware that the "user" will be continually gaining experience.

Categorisation of users has been useful, but the emphasis has always been on the user adapting him/herself to the system rather than vice versa. The focus here is on the possibility of categorising users still further in an attempt to personalise the system to the user.

The main distinction made when referring to adaptive interfaces is between adaption by user (Figure 1.1) and self-adaption (Figure 1.2). The former refers to a system which the user may change according to his/her preferences. A self-adaptive interface is controlled by the system and may be programmed to accept information about the user. The intelligent front end uses this a priori information to make decisions about how to interact with the user. Further information can be gained through the user's performance which may cause the intelligent front end to make modifications to the appearance of the interface. When I speak about 'the interface', I am referring to all aspects as defined above.

The general emphasis in this thesis is that the interface should be constructed from a user-centred viewpoint. One type of information that can be used for making decisions about the appearance of the interface is the user's cognitive and learning styles and other personality

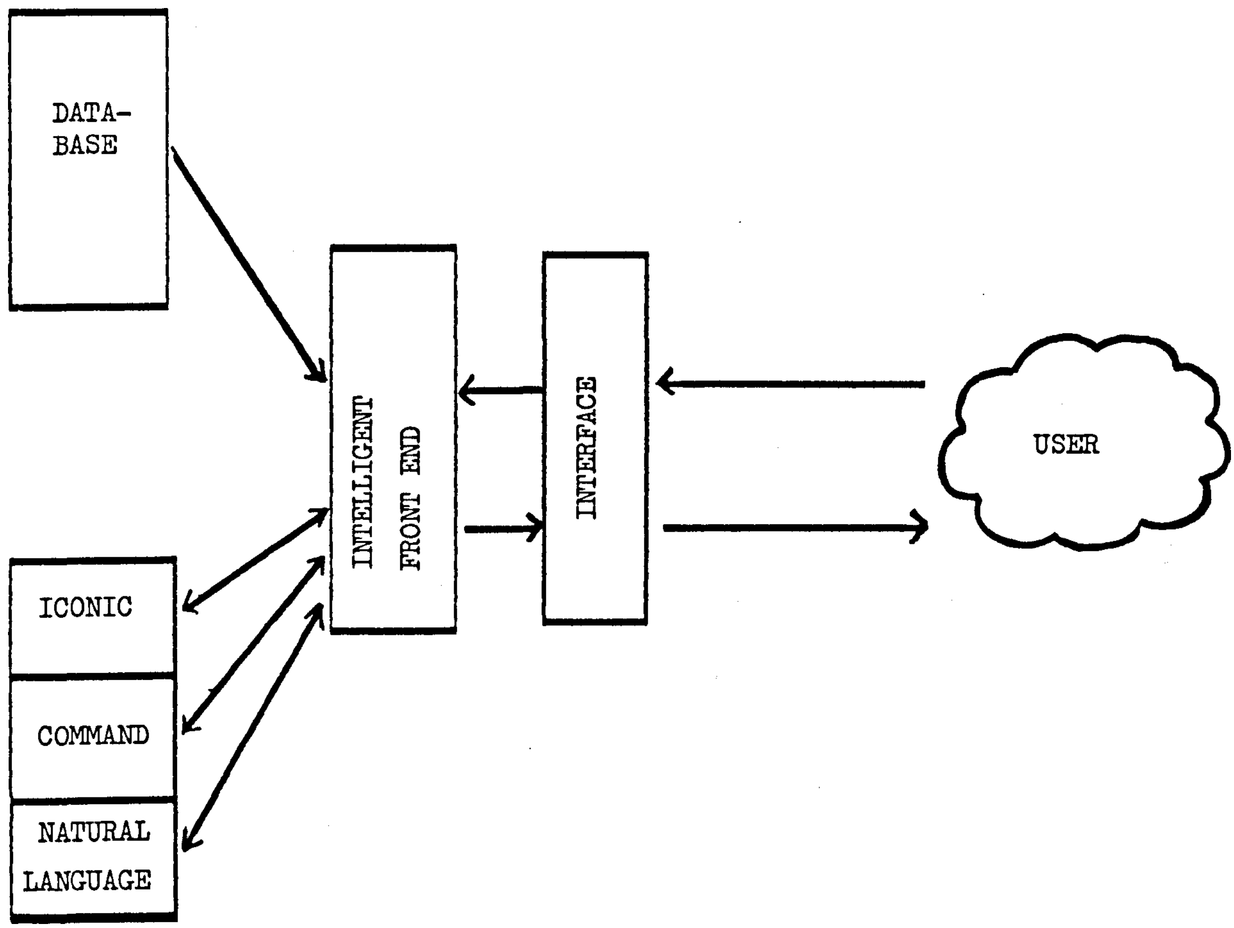


Figure 1.1
Schematic Diagram of a User-controlled Adaptive Interface

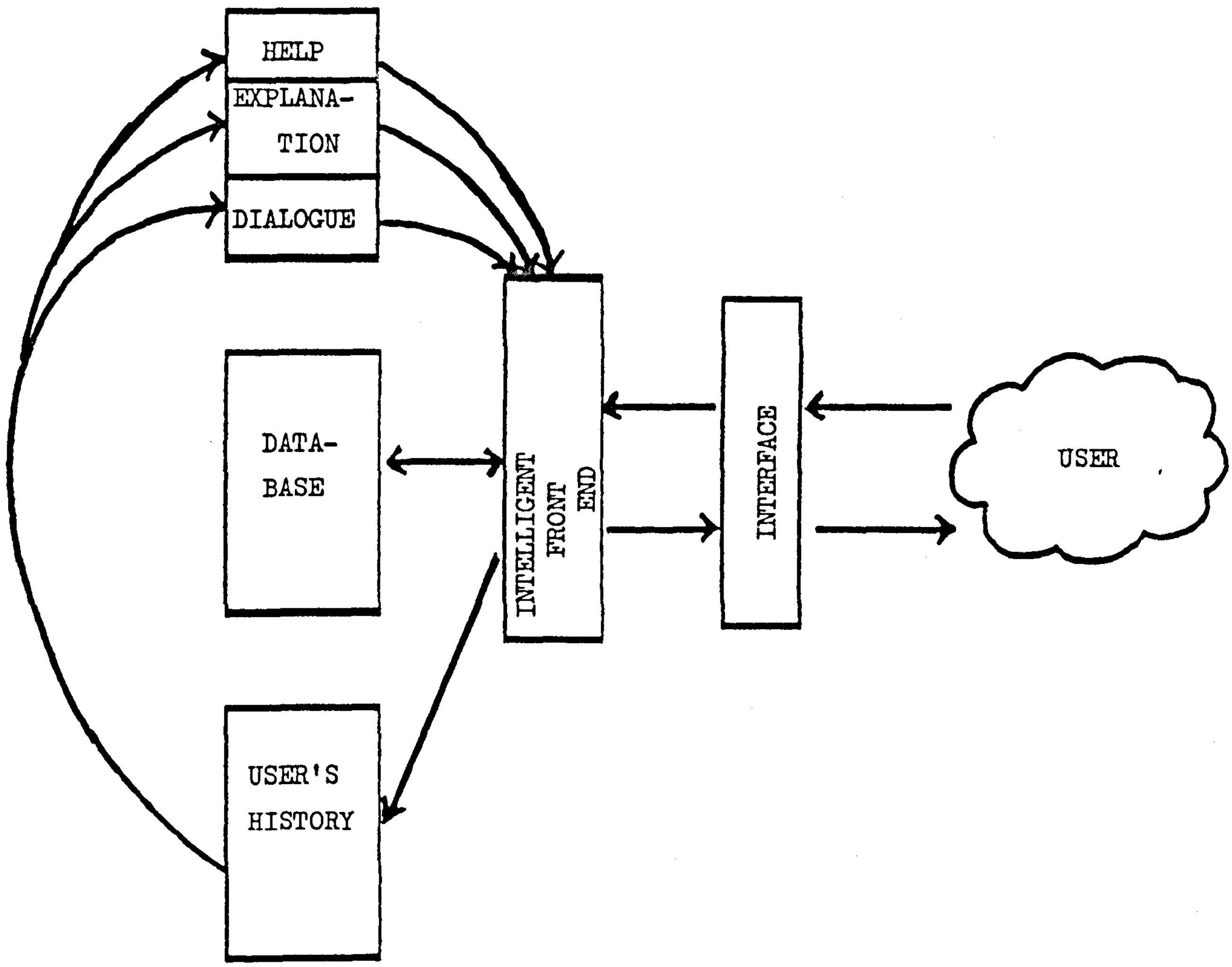


Figure 1.2

Schematic Diagram of a Self-adaptive Interface characteristics. In this thesis, these characteristics are referred to under the general heading of "personal style".

Cognitive style refers to the way a person perceives a problem while learning style refers to the way a person prefers to learn in a learning or problem solving environment. Such information can give some idea of what the user will find difficult when interacting with a computer and how to prevent these difficulties. Personality factors such as introversion-extraversion are also thought to affect the way people perform on problem-solving tasks (e.g., Allsop and Eysenck, 1974) and therefore affect the type of environment that would best suit him/her.

In sum, the types of difficulties experienced by particular users may be decreased by providing a flexible interface which is compatible with the user's personal style. For example, a user who possesses a number of characteristics to suggest that s/he may experience difficulties can be given an interface which uses a graphical form of dialogue or menu-driven system. Help, explanation and learning facilities can be extensive and written in the user's learning style. In addition, a faster rate of feedback can be implemented, as well as the use of colour and graphics, to decrease processing and memory loads.

We need to evaluate the usefulness of knowing about a user's personal style. In the domain of Computer Aided Learning (CAL), the concept of personalising systems is important. Research by Pask and Scott (1972) has demonstrated the usefulness of giving students material written in the same learning style as that of the student. It would seem useful for intelligent tutoring systems to present learning material in a way most compatible with the student's learning style. Level of explanation and support could also be modified accordingly. Since most computing systems will also have a tutoring implication associated with their use the findings of Pask and Scott may have wider import for a range of applications.

The usefulness of self-adaptive interfaces or intelligent front ends to expert systems is debatable. For example, if an expert system is also to operate as an intelligent tutoring system then the points identified for CAL apply. Their usefulness rests on the type of environment in which the expert system operates and its frequency of use. Some expert systems will tend to be used sporadically by many people who possess varying amounts of knowledge not just of computers but of the actual knowledge in the domain. If the knowledge domain is medical, requiring quick diagnoses, then it may be more economical to choose other ways of categorising the user. For example, users could be categorised as expert, familiar and novice and then categorised further in terms of how much knowledge each particular user possesses. Modification of communication mode, language used (e.g. lay or technical) and support and explanation facilities could be made accordingly.

If, on the other hand, the system was used frequently by the same set of people, then it would be feasible to build up personal files of each user. The personal file could possess information about the user's personal style, which would in turn be used to modify the interface for that person. Performance details could be collected and later used to make further modifications to the interface.

In the domain of word processing, it may be possible to give users on-line learning materials and 'help' facilities compatible with the user's personal style. The user's level of experience will also be important as one would expect it to interact with cognitive and learning styles and personality. One possible application would be to provide the user with an interface to a word processor appropriate to his/her level of experience and then to vary the amount of support within the interaction in relation to the user's personal style. This is likely to be particularly important when the word processor is being used by a document author, rather than a clerical transcriber of text already written.

1.1.1 Types of Support

The support a user experiences during an interaction is encompassed in the entire interface. The various types of dialogue with which computer and user communicate offer varying amounts of support. Iconic based or graphical systems, in conjunction with pointing devices, tend to offer the greatest amount of support because the user spends less time typing in commands at the keyboard and more time manipulating icons which produce actions. Menu-driven systems also offer support by prompting the user rather than vice versa. Finally, command languages give full control to the user and typically offer very little immediate support.

1.1.2 Help Facilities

Help facilities as opposed to generalised support facilities are specifically designed to give information about how the apparatus operates. On-line help facilities can be varied according to the user's level of experience. Help facilities include explanation and the extent of explanation and even variations in the complexity or style of the language employed in the help facility. It may be useful to provide extensive and complex explanation which is compatible with the user's personal style, but for short explanation, it would probably not be economical to take such an approach. Extensive explanation is likely to be required for computer aided learning, expert systems and word processing learning material.

1.1.3 Colour

The employment of colour on computer screens, in particular in relation to video games and Prestel type advertising, is common. It has been used to a lesser extent in word processing packages. The main use of colour has been to improve the aesthetic quality and to make the screen more interesting. Colour, however, does have more functional uses. For example, colour coding of targets aid their search from the background over other achromatic coding. This type of colour coding has most widely been utilised in aeronautical applications such as the pilot's cockpit and computerised maps (e.g. Luder and Barber, 1984).

Colour coding can also be envisaged as a type of support which may give information about a user's location, for example. The common problem of user's "getting lost" in hierarchies of menus can be eased by colour coded screens. But this still entails the user having to make the correct associations between the colours and their respective location. This type of application employs non-qualitative colour.

Colour can also be utilised in a qualitative manner, although this is rarely used in a complex way. Qualitative colour refers to colours which possess inherent meanings which are therefore not newly learned. When these colours are used they can induce people to make qualitative judgements; for example, green to represent "proceed" and red to represent "stop".

1.2 The Framework

The idea of a self-adaptive interface is that it should appear different to many users of the same system. It presents an interface which seems to "think" and "communicate" in the same way as the user.

Figure 1.3 is a schematic diagram of how adaption of the interface might occur. Basically, the user's level of experience, and both cognitive and learning styles and personality characteristics are used to construct a personal file which can be later modified by the user's actual performance. From these details, the extent of the support facilities are established and the mode of communication chosen. The intelligent front end then modifies the interface accordingly. The user's performance during the interactions will change as the user gains more experience. These performance details reside in the user's history and are used to modify his/her personal file which in turn may modify support facilities and the mode of communication.

Cognitive and learning styles and personality characteristics are relatively static factors and are therefore likely to be useful in establishing the personal file to define the appearance of the interface. But what is the function of static factors after the initial stage?

m=modify
.... modifiable components
____ non-modifiable components

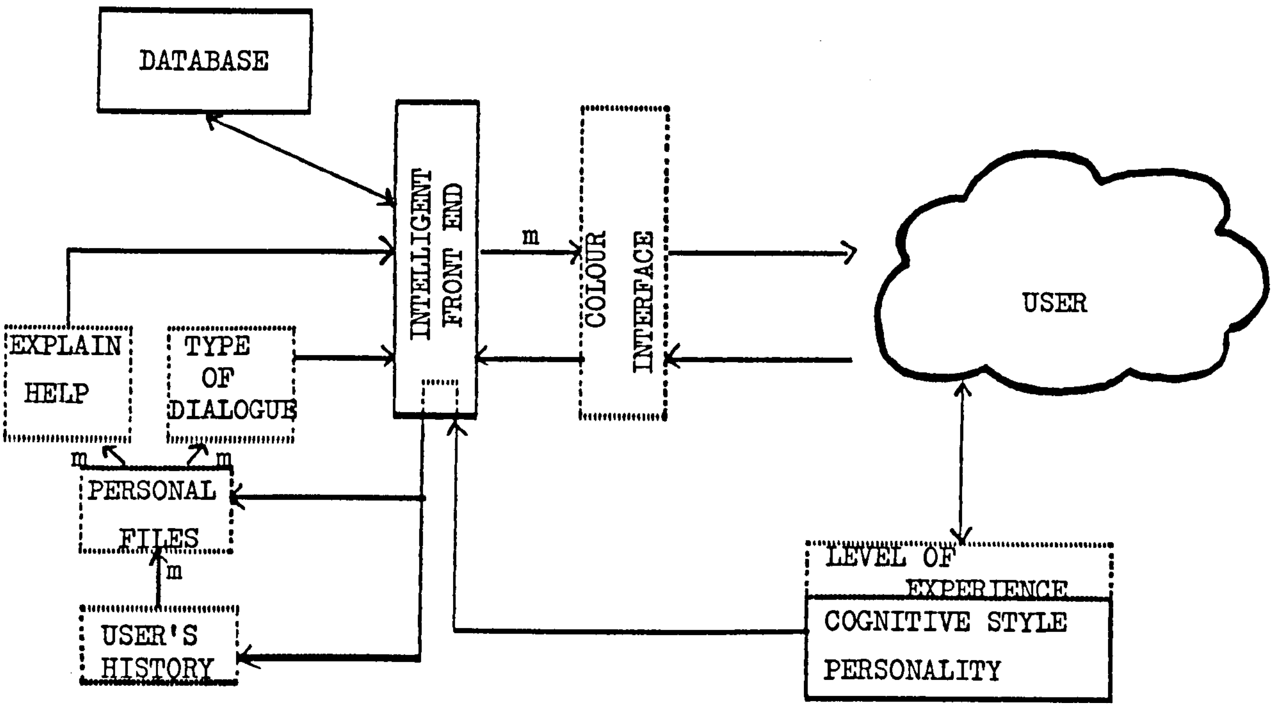


Figure 1.3

Schematic Diagram of a Self-adaptive Interface from a User-centred Viewpoint

More dynamic factors such as performance will be important in modifying the interface as performance tends to change with practice. On the other hand, cognitive and learning styles and personality are important at later stages because the user will still experience the same types of difficulties irrespective of experience, precisely because these factors do not change, or change very slowly over the lifespan. The overall result is continual modification of the interface as performance increases but concurrently ensuring that all support facilities are compatible with the user's personal style.

This chapter has introduced the concepts to be further developed in the thesis and demonstrated how these concepts fit together within a framework. The next chapter consists of a literature review of the area of research, beginning with a review of work on adaptive interfaces and then introducing the concepts of cognitive and learning styles and personality characteristics into human computer interaction research. Upon this theoretical basis lies the set of experimental studies reported in Chapters 3 to 5 and 7 to 10.

Chapter 2 Adaptive Interfaces

2.1 Introduction

This chapter is divided into two sections. The first section provides a literature review of adaptive interfaces from a human factors viewpoint, while the second section explores the relationship between personal style and human computer interaction and its relevance to adaptive interfaces.

During the last decade many researchers investigated the notion of adaptive interfaces including their desirability. Adaption may be of two types:
(a) adaption by the user allows him/her to <u>adapt</u> or <u>tailor</u> the interface to the needs of a particular task e.g. changing the mode of communication. User adaption recognises the fact that user performance varies within the individual over a period of time and that there are differences in performance across individuals. For example, novice and expert users have different requirements.

A study by Klemmer and Lockhead (1962) found that, among a group of skilled data input operators, variations in speed were in the region of 2:1 and errors a staggering 10:1. Sackman (1970) provides further support for wide user variations in a study which recorded subjects' number of moves and amount of time to complete problem solving tasks. Subjects varied in the region of 15:1.

This evidence indicates that user variation needs to be taken into account in interface design. The match between dialogue and explanation or help facilities and user type is crucial to the success of an interaction. In general, computer-initiated forms of dialogue, such as form-filling and natural language, menus and desktop workstations, are thought to be more suitable to tyros than those more experienced in computer use. Conversely, command language or direct mode, which is a user-initiated dialogue, is thought to be more acceptable to expert computer users.

(b) Automatic adaption or self adaption is when the system or intelligent front end changes the dialogue strategy and offers help according to how it "perceives" the user's progress. Thus, it automatically adapts to the user in response to the dialogue history. Automatic adaption is complex, first, because the computer has no previous knowledge of the user's performance (a training phase is needed); and, second, because the interaction may be degraded by the fact that there will be low stability or consistency within an interaction. That is, the user and the computer will continually attempt to model each other (Gaines and Shaw, 1983).

A third type often suggested is adaption by the system designer; that is, the system designer alters the interface to suit the experiences of the user (Edmonds, 1981).

The power of an adaptive interface lies in its suitability for all users, that is to create "a system that actually appears different to different users" (Rich, 1983).

One of the consequences of the widening gap between the availability of new technology and the manner in which the new technology is implemented is that the types of machines available on the market are to a greater or lesser extent, unusable. The factors affecting usability stem from rigorous task and user analyses. In other words, they stem from the degree of compatibility between the type of user and the types of tasks the user will need to perform and the ability of the machine to allow the task to be performed easily. To build a system with these prerequisites is not an easy task because potential users are not always able to specify what it is they require (see Edmonds, 1974; Kidd, 1986). In addition, the requirements of the user change over time. The fact that the novice user soon becomes an experienced user changes aspects of the task performance. There is empirical evidence to suggest that experienced users demand a faster response and more control over the system than tyros (Benyon, 1985). Thus, the problem here is two fold: interface designers need to take into account both individual differences in terms of their needs as users e.g. secretary, businessman, computer programmer, and individual changes in experience and familiarity over time.

2.1.1 Building the User Model

Rich (1983) has pointed out that, in general, those computers which provide some amount of tailorability are limited to attempting to model an abstract typical person rather than particular characteristics of the individual. This type of user modelling is not adequate for adaptive user interfaces not least because it confounds within- and between-group variability. Instead, she has proposed that a more useful way of modelling the user would be a system which possesses a collection of models of individual users.

Similarly, models which are inferred by the system, rather than explicitly specified at the design stage, will permit greater adaptability, as will models of short term user characteristics more than long term user characteristics.

The techniques Rich suggests for building user models are: identification of a user's typical vocabulary and concept usage, gauging the types of responses with which the user seems satisfied, and employment of stereotypes to generate many facts from a few facts. One way of inferring individual facts is through observation of the user interacting with the interface. For example, the pattern of commands used will reflect aspects of the user's understanding. However, performance and usage does not always accurately reflect competence and understanding. This type of user modelling is found mostly in CAL systems in which it is important to gauge the progress of the student in order to be of any educational use. The main problem arises from the collection of all the necessary information about the user and his/her performance within a relatively short period of time. Stereotypes or "scripts" (see Minsky, 1975; Schank and Abelson, 1977) permit many facts to be inferred from a single or small set of facts. A stereotype represents a collection of traits which can be represented by the system as a set of facets filled with values. The facets of the stereotypes used by the system should correspond to the facets of the user models built by the system.

Some work has been conducted on the cognitive style of the user (e.g., Fowler, Macauley and Fowler, 1986; Macauley and Norman, 1985). Van Muylwijk, Van der Veer and Waern (1983) have argued that cognitive style and personality factors are characteristics of the user relevant to the learning phase and can be expected to remain stable over an interaction. These characteristics can be built into a stereotype which can be invoked for particular users (see also Benyon, Innocent, Shergill and Murray, 1986; this chapter).

Geiselman and Samet's (1982) work supports the usefulness of adaption of the interface. They found personalisation of message formats to be useful and led to improved subjective organisation of the intelligence data. That is, more information is comprehended when material is schematised in a manner consistent with the learner's organisational view of the topic presented. They propose that different readers of messages can have different output formats to suit their needs and this will increase the rate of information assimilation.

The implication is that people should be able to work in a way that best suits his/her way of thinking. For example, some computer users will prefer graphical interaction, others textual dialogue, and others, a mixture of both.

One problem which may be encountered in attempting to personalise the interface is raised by Stewart (1974) who writes "it is important that the formal computer procedures do not prevent the user from changing his representation of the problem or the task environment necessary to reach the best solution". Karat (1986) provides support for the above statement. He found that users are sensitive to only minor changes in the interface. Problems arise because computer users find it difficult to transfer from one type of communication mode to another or from one system to another. More problematic are cases where different modes of communication are combined. For example, there may be particular tasks which are easier to perform using graphical techniques and other tasks which are more easily performed using command mode. However, by default, computer-initiated and user-initiated forms of dialogue fall at opposite ends of the continuum and probably do not mix easily.

In order to create adaptive interfaces we need to be able to modify the interface easily. This, in turn, requires easy-to-use tools (e.g., see Edmonds and Guest, 1978) which must recognise the fact that part of user behaviour will be unpredictable. Again the implication is that designers need sophisticated models of users which permit accurate evaluation of user information and behaviour in addition to the ability to specify and represent this knowledge formally (Innocent, 1982).

2.1.2 Adaption By User

A large area of research in human computer interaction has focussed on type of user and the dialogue most appropriate for a particular user. The research has provided information for work on adaption. The methodologies employed in the studies differ widely, using a variety of materials and procedures as well as using different criteria for subject selection. The main concern of the studies is to investigate different types of users - novices and experts - while working on particular tasks (mainly text editing tasks) and recording performance measures such as moves to complete a task and errors. Questionnaires have also been popular as a way of establishing the level of users' experience. The weakness of such techniques is the lack of theoretical foundations for classifying user types. For example, what is a 'novice user' and what is an 'expert user'? Within the research literature the difference is usually defined opportunistically, by the comparative characteristics of available groups. Thus, it is almost impossible to derive firm descriptions of the characteristics of experts and novices. Additionally, attitude and motivation seem to contribute to learning to use computers (see Breakwell, Fife-Shaw, Lee and Spencer, 1986).

Folley and Williges (1982) have developed a method for determining novice and expert users' models of interactive text editing by empirically analysing patterns of command language use. Twenty novice users (no interactive computer experience) and twenty expert users (median of years of experience was 3 years and median of hours spent working on an interactive computer terminal was 25 hours) performed three text editing tasks. The text editor was created by asking 12 novice and 12 expert computer users to indicate on a questionnaire which command they believed to be the most appropriate description of a particular action. Subjects were also asked to list the commands they would use for each text editing task in the correct order.

The findings suggested that the experts' user model was essentially the same as the novices' user model set of commands with the addition of some more powerful commands. This particular study investigated the use of command mode across user types. However, it is already commonly believed that novice users benefit from graphical environments and in reality novice users rarely interact with command mode.

Walther and O'Neil (1974) compared two text editors - one flexible and the other inflexible. Experienced users were found to work more quickly with the flexible system than the inflexible system whereas the opposite was found for novice computer users.

More recent work has also provided evidence that novices prefer structured system-led interactions. For example, Ogden and Boyle (1982) compared three language modes - command, form fill-in and a hybrid of these two where a subject chose a command plus a parameter. Novice computer users preferred a computer-initiated form of dialogue: form fill-in. No follow-up study was conducted so it is not possible to locate when a potential switching of mode preference may have occurred for the novices.

Gilfoil (1982) supports Ogden and Boyle's findings. They studied novice computer users to investigate the transition between two modes of communication - menu-driven and command-driven languages. The novice users performed six computer tasks over one month, totalling 20 sessions, in which they were allowed to choose between using either menu-driven and command-driven languages at the beginning of each session. Cognitive, affective and performance measures were recorded. Menu-driven languages were found to be more appropriate for these novice computer users with transition to command-driven after approximately 16-20 hours of computer use.

Hiltz and Turoff (1982) conducted a similar study with the aim of observing the use of different modes of communication. Their study differed from the others in the length of the observation period (18 months). Very few studies have been longitudinal in nature. Differences in the length of the studies renders it difficult to compare and assess the results of studies investigating the novice to expert learning curve.

In the Hiltz and Turoff study, data were collected by questionnaire, before use, at 3-6 months and after 18 months of use, in addition to employing monitor statistics, to register the amount and use of different interface modes. The modes offered were menu-selection, command-driven, answer-ahead and command streams and self-defined commands. Hiltz and Turoff expected learning to follow a progressive curve with most users starting with long menus through to short menus, to answer aheads and commands and then finally defining their own commands. This was partially supported by their findings but in addition they found that, although the menu was preferred by new users, its use did not completely cease. In particular, users regressed to menus when traversing new or unfamiliar parts of the system, or alternatively after a break in using the system. It appears that the more experience a user gains the more likely s/he will use command modes, but the other modes are never lost entirely. The findings suggest that there is a need for mixed modes of dialogue and that they can be potentially useful because the user is permitted to dynamically change the type of interaction and therefore be in control. Very few systems at present combine mixed modes of dialogue although some word processors do provide faster ways to achieve an action.

Potosnak (1986) has looked specifically at preference of communication modes by classes of users. She has criticised earlier studies for their lack of theoretical bases in the classification of users participating in the studies. The majority of studies have used one single variable to group the subjects, most commonly in terms of amount of computer experience and other subjective variables. Potosnak used cluster analysis on 45 variables for 481 potential computer users. She found that all her subjects preferred to use the prompt mode over command and form-filling modes behaved accordingly. The main difference between groups was that the expert group switched modes more often than other groups and performed more actions using the command mode. Potosnak concludes that her research refutes the hypothesis that different users prefer different interface modes. However, the study does not include any results which might suggest that all users perform better using the prompt mode.

Other research has demonstrated that the frequency of use and length of the interval between usage also affects the time of transition between two language modes. For example, Maskery (1985) has conducted a study on the progress of novice users when interacting with a computer-based tool, via an adaptive interface to a statistical and graph-plotting package, called Dialog. He recorded user difficulties and the overall reaction of the user to the adaptive interface. The adaptiveness was based on three levels of dialogue style (see Miller and Thomas, 1977). These were: (a) system leads - user has forced choice, (b) system leads - user has free choice and (c) user leads - user has free choice. After the user completed a task at level one, s/he automatically went onto the next level. No switching of modes was possible after a level had begun. Frequency of use could be occasional (2 sessions at intervals of 6 weeks), intermittent (6 sessions at intervals of one week), frequent then occasional (5 sessions at intervals of one day and then 1 session after 5 weeks). Maskery found that the change from system-led to user-led dialogues resulted in an increase in errors, performance times and calls for help. Often the subjects were not expecting the level change or were unaware of what had happened to cause the change. Subjects who used the tool on a weekly basis demonstrated superior performance to the others. Typically, these subjects were not as distressed as other subjects when an error had been made, showed a greater decrease in refamiliarisation times and changed to level three earlier than subjects using the tool daily or occasionally.

More recent work (e.g., Hanne, Hoepelman and Fahnrich, 1986) has attempted to combine different media in an adaptive interface which will allow the user to choose his/her preferred type of input in addition to switching between media during the interaction. Hiltz and Turoff (1982) have provided one good reason for doing this; that is, the system would allow users who are traversing new parts of the system or alternatively those who use the system intermittently to regress to more familiar ways of achieving actions. On the other hand, very few studies have tackled the problem of mixed modes of communication. Thus, we do not know what effect mixed modes will have on the users. An inhibitory effect on performance may occur as a result of mixing different internal representations and altering the solution path. One could expect transfer difficulties similar to those experienced when transferring from one system to another.

Motivation has also been found to affect performance. Harris (1977) found that those subjects who really needed to know the information in the ROBOT natural language query system database were much more tolerant than those subjects who were not particularly interested in the answers given. Similarly, negative attitudes decrease performance (Walther and O'Neil, 1974).

In sum, novice computer users have been found to prefer system-led dialogues which give little room for flexibility. Their command repertoire is essentially the same as the more experienced but is less extensive.

However, the interaction between communication modes and user type is not clear-cut: there appear to be many other factors intertwined such as attitude, motivation and frequency of use.

One of the limitations of the studies reviewed here lies in the lack of a theoretical basis for the studies. Potosnak (1986) has pointed out that the result of this is that comparison of the findings from different studies is difficult. The problem of adaption cannot be satisfactorily dealt with until an adequate and standard classification of user types and of factors which may or may not be desirable to modify is available.

Another limitation is that most of the studies are short-term. This has prevented any quantitative meaning being assigned to the variables "novice" and "expert". Again, lack of well-defined terminology does not permit one to draw general conclusions from the studies.
Very few of the studies mentioned above tackle the problem of mixed modes of communication. More research needs to take place concerning the advantages and disadvantages of using mixed modes of dialogue.

Finally, research needs to focus on creating uniform terminology which is based on more than one or two variables (e.g. computer experience). This would allow a small number of categories to be identified and hence provide a theoretical basis for the common research. In addition, research aimed specifically at understanding the user's model in terms of the schematised organisation of the information to be handled would be useful.

2.1.3 Self-Adaption

The principal requirement of self-adaptive systems is that they should be able to maintain an adequate and up-to-date model of the user throughout the interaction. Some aspects of the user model are stable (e.g., user characteristics, personality, learning and cognitive styles) and some are unstable (e.g., the current task and user experience). User experience in particular has a major effect on dialogue because the dialogue indicates the intentions of the user and reflects the user's understanding which in turn influences the system's responses and the amount of guidance given through the dialogue network.

Edmonds (1986) provides a categorisation of adaptive features for the end user which may be evoked by the user, prompted by the system or automatic adaption by the system. Common user errors such as misspellings can be automatically corrected without the user's awareness but this introduces the risk of incorrect correction, especially for proper nouns.

User characteristics include motor co-ordination and perceptual skills which tend to be stable characteristics and so can be automatically evoked at the beginning of an interaction. User performance tends to be more variable dependent on the length of time spent on the system. Fatigue will affect the form of help or explanation required and the preferred structure of the dialogue.

User goals are difficult to model because they are task related. Prompted adaption or adaption by the user whereby the user states explicitly his/her goals may be possible but there is no technique for the system to adequately model these goals and thereby adapt itself accordingly. In the domain of Artificial Intelligence there has been some attempt to use plans and goals (see Allen and Perrault, 1981; Cohen and Perrault 1979) but as yet these systems are extremely primitive in their capabilities.

The information environment in which the task is to be achieved is also very difficult to adapt to either automatically by the system or by prompting the user.

Two research questions that Edmonds raises are, first, how would users react to such adaptivity? They may find it daunting and, second, how can systems which will provide such facilities be built?

The first question raises issues of desirability and some of the studies to be reviewed touch on this factor. It may be that extensive adaptability or flexibility of the system will leave the novice end user confused because s/he will not be able to form a stable model of the system's functions because they are constantly changing.

Benyon's (1985) MONITOR system is derived from Innocent's (1982) conceptual model of a self-adaptive user interface (SAUI). The conceptual design of MONITOR incorporates Rich's (1979, 1983) approach to creating a user model by referring to a number of stereotypes. MONITOR possesses a number of stereotyped dialogues or scripts which can be adapted to the user. Carroll and Thomas' (1982) metaphors have also been used as part of the conceptual design. Metaphors are useful for representing stereotypes of different tasks to be accomplished by the user. The dialogue itself is represented by a transition network where the nodes represent points in the dialogue and the arcs possible responses.

MONITOR has been developed as a prototype of an expert system in the domain of computer-assisted training. The architecture contains both procedural and declarative knowledge. The main components are: knowledge of control which selects the appropriate task and dialogue; the knowledge base which maintains the user model (details of user, tasks and performance) via scripts; and the database which contains declarative knowledge.

Benyon et al (1986) have further extended this work on CAL projects. This system is based on Rich's (1982) GRUNDY. The user model consists of a user profile and a user history. The former contains personal knowledge and stereotypical features inferred from the individual's college course and personal questionnaire data and the latter contains user history records, all user responses and their path through the dialogue network. Parameters such as prior knowledge, learning style, heuristic competence, intellectual ability, motivation to learn, interest level and cognitive style are set up as stable components of the stereotype and are activated at the beginning of a tutorial session. The user's knowledge of the particular concepts being taught are assessed. Each concept to be learned is assigned a confidence value which reflects the user's understanding. All this information is used to decide how much guidance should be given through the dialogue network.

Cooper (1986) and colleagues have also been working on tools to assist the development of adaptive user interfaces. Their adaptive interface has been specifically built to interface with Telecom Gold. The main aims are to support different levels of skill across tasks and users, to increase the task flexibility of the system and to allow users to draw on prior experience (of mail systems). The interface can adapt in three ways: by changing the level of guidance (change in use of prompts and feedback), context (permitting a short term change in user's goal and thereby adapting to the user's conceptual model) and analogous mail (ability to recognise commands from other mail systems).

The user interacts with the adaptive interface via a VT100 terminal. The dialogue controller executes task-sized fragments of user dialogue from which the user model selects particular task fragments. From this the user model is able to infer the user's task and determine the user's skills on that task. On the basis of this determination, the amount of guidance may be increased or decreased. The application expert checks the user's task specification for semantic errors. If the specification is accepted then it is translated into Telecom Gold commands in order for the interface to interact with the system. In turn the dialogue controller selects an appropriate response to return to the user. If the specification is not accepted then the application expert must use its knowledge to decide on a recovery strategy.

The adaptive interface was evaluated against its non-adaptive counterpart using four members of the public, three of which were experienced operators. Subjects performed tasks (deleting, filing, forwarding letters) in three sessions which were recorded on video and an on-line log, as well as a questionnaire which elicited their feelings about the interface. Cooper reports that the user model was able to detect approximately 40% of user difficulties but users found only 19% of the detected errors useful. Further evaluation of the interface was conducted by Cooper and Hockley (1986) using a total of nine subjects - six of whom used the adaptive system and three the system in its default mode. They reported that, in general, both performance on and opinion of the adaptive system was significantly lower than for the non-adaptive system. Comparison of the changes in the user model and user's perception of the difficulties again showed that the adaptive interface could detect the nature of around 40% of the difficulties that occurred. It was found that the actual changes that the system was prescribing, in order to recover, were inadequate with subjects reporting just 7% of them as being helpful. The increase in errors led to increasing verbosity of the explanations given. They concluded that providing increased amounts of pre-stored text did not help the user in clarifying his/her misunderstanding and a finer-grained user model is required. This work also demonstrates the importance of prototyping and evaluative work. Literature has identified the components thought important and necessary in constructing a usable interface but, in practice, an adaptive interface that possesses sophistication and extensiveness is extremely difficult to build.

Greenberg and Witten (1985) have created an adaptive telephone based on the frequency and recency of telephone calls. The adaptive telephone directory was assessed against a non-adaptive telephone directory. The two directories differed only in how the user model was created. The former was dynamic and changed with the calls made while the second ignored the user's history of calls. The results of their study demonstrated that the adaptive telephone directory was more efficient in terms of trial time taken, errors made, scanning time and preference.

Alty (1985) has developed an interactive dialogue system called CONNECT which uses a combination of transition networks and production rules to guide the dialogue. It also uses path algebras to assist the production system in making topological decisions about the network as well as performing an analysis on the network. The analysis of the network leads to the construction of adjacency matrices which contain values associated with the labels for the network arcs which may indicate labels for alternative arcs and arc sequences (for details of the analyses refer to Alty, 1985). The networks are specified at the outstart and modification occurs only through the opening and closing of arcs. The dialogue network consists of parallel networks, each of which shows a different dialogue representation to the user: those arcs which change the representation of the interface can be opened and closed. Adaption is controlled by the production rules via switching matrices which permit the switching between one network to another. Adaption of the nets occur automatically as a result of past arc transitions and other user information.

In summary, self-adaptive interfaces are still at a primitive stage. It is generally agreed that the architecture of such systems should separate the dialogue from task components. But there are also problems of representation: the structure of the components must be modifiable to suit particular users. Thus, some of the knowledge is uncertain or "fuzzy", in particular knowledge about expectations. A second prerequisite is that the knowledge should be represented in such a way as to allow a system to respond in a relatively short period.

Many of the systems mentioned above are more conceptual rather than working implementations. With the exception of Telecom Gold those systems which have been implemented have not provided evaluative reports and without these they are difficult to improve.

Another important question raised by a number of researchers (e.g., Edmonds, 1981) is the nature of users' reactions to a self-adaptive interface. A user may feel confused by his/her changing model of what is going on in the interaction. The novice user tends to have a very blinkered view in his/her approach to computer use and often needs to understand specific responses in terms of specific actions. Thus, if the response of the computer to the same action varies the user may find it difficult to build a stable model of the interaction and one would expect learning to be slower when using an adaptive interface as opposed to a non-adaptive interface.

The types of adaption which have been described so far have been concerned with behavioural measures such as performance and occasionally preference for a particular input media. Another way of approaching adaption which is described below concerns the user's cognitive and learning styles and personality and through observing how these factors might affect performance on a computer system. This type of observation exceeds that of investigations of categories of users such as novice and expert because it deals with the individual per se and accepts that there are many inter-individual differences not captured by such broad categories. Furthermore, as described in chapter 1, personal style can be integrated in a self-adaptive system to make the interface appear different to different users.

2.2 The Relevance of Cognitive and Learning Styles and Personality factors to HCI

The relevance of the personal style of the user to the ability to use computers effectively has become a major issue in the design and creation of adaptive interfaces. That these factors tend to be both quantifiable and consistent within the individual provides a basis for the classification of individual users according to the manner in which they might approach a particular problem; that is, the way the individual structures and reorganises incoming information when learning new material as well as their ability to handle situations involving high amounts of anxiety. In turn, this may suggest the type of environment in which a user will perform at an optimum.

2.2.1 Cognitive and Learning Styles

The majority of the work associated with cognitive and learning styles has been oriented towards education and assessing children's approaches to learning on the basis of available tests. Some of the tests developed to measure cognitive styles were designed specifically for testing children, such as the Matching Familiar Figures Test, while others were designed for testing adults and later adapted for children, for example, the Embedded Figures Test (including the group version). Such tests have only recently been adopted in the area of adaptive interface design. The general assumption is that novice computer users tend to manifest particular cognitive and learning styles while interacting with the computer and that these may be classified and used as a basis for assigning appropriate dialogue styles. Learning to use a computer can be seen as a problem-solving task where the user has to assimilate incoming information into a structure - a mental model - of how the computer works. The way a user attempts to do this will depend largely on his/her cognitive and learning styles.

2.2.1.1 Field Independence-dependence

The most widely used test for categorising individuals on the field independent-dependent dimension is the Embedded Figures Test (EFT), or the group version (GEFT), devised by Witkin, Ottman, Raskin and Karp (1971). Witkin and Goodenough (1981) proposed that the degree of psychological differentiation displayed determines an individual's autonomous functioning or level of field independence. The EFT measures the ability to differentiate item from context: those individuals scoring high on the EFT are labelled field independent; that is, they are internally driven. They are able to analyse incoming information and actively restructure this information in relation to the demands of the situation or their personal needs. Field dependence refers to a cognitive style possessed by people who are not self-reliant and are externally driven by context or the situation. They tend to be "passive" recipients of information.

A number of researchers have suggested that field dependence-independence is important for the way users construct their mental models of how a system works and how to handle tasks. Van der Veer et al (1985) suggest that field dependents lack the capacity to use mental models of a familiar system as an analogy for acquiring an understanding of a new system.

Fowler et al (1986) investigated the relationship between field independence-dependence and performance on different dialogue structures. The data were taken from a study by Macauley and Norman (1985) who tested clerical staff on text-editing and retrieval tasks, all of whom possessed little or no experience of computer use. The complexity of the command language differed on two dimensions: one language had a linear structure (command and arguments required in a single user response) and the other used a substructure form (a set of prompts for the arguments required by any one command). Relationships between GEFT scores and performance measures (thinking time, doing time, number of usage errors, number of typographical errors, number of errors corrected, number of errors left uncorrected, number of times the help facility was used) were established using correlational matrices.

Fowler et al found that field independent users were able to develop their mental model of the task requirements before undertaking the task and then were able to refine the model through experience. They suggest that extensive detailing of system operations would benefit this class of users. No difference in performance was observed between the different language structures for field independent users. On the other hand, field dependent users learned on a trial and error schedule and their models of the task developed more slowly. Fowler et al suggest that these users need basic operation information and then to develop their model through experience. Field dependent users performed better on a substructure form of dialogue which prompted the response and less well on a linear form of dialogue. The degree of difference in performance between the field independents and the field dependents was dramatically reduced on the second block of trials, suggesting that the field dependents had begun to catch up the other users in performance. If Fowler et al are correct in saying that field dependents performed using a trial and error schedule, it is not clear whether field dependents possessed a similar level of understanding as field independents. From what is known about how field dependents and independents perform on computer tasks, field dependents should have less understanding of the computer's operations.

These results suggest that dialogue structures should be compatible with a user's cognitive style. For example, field dependents need a supportive environment in which to interact, including a large amount of direction at the early stages, and eventually are able to progress to a less supportive environment. On the other hand, some field independents may prefer to begin with a computer-initiated language mode but switch to a user-initiated or less supportive language mode at an earlier stage while others may prefer a user-initiated language from the beginning.

In the domain of Management Information Systems (MIS) Bariff and Lusk (1977) and Lusk and Kersnick (1979) found that field dependents and independents have a differential preference for data presentation. Field independents prefer detailed, aggregate quantitative reports while field dependents prefer to study the raw data. Benbasat and Dexter (1979) arrive at similar conclusions. They found that field dependents have better performance on a database inquiry system which showed raw data of up to 50 days while field independents requested more reports when given the disaggregated information of the database inquiry system. Additionally, where a mismatch between field dependence and report format existed, more reports were requested.

2.2.1.2 Reflectivity-impulsivity

Reflectivity-impulsivity refers to the tendency to reflect on a problem before choosing one response where several options are available and where there exists some doubt about the correct response choice.

The Matching Familiar Figures Test (MFFT, Kagan, Rosman, Day, Albert and Phillipa, 1964) was originally developed for children and then revised for adults. The test involves the simultaneous presentation of a figure (e.g., boat, telephone, with 4, 6, or 8 facsimiles differing in one or more details). The time the individual takes to make the choice and the error rate are recorded. This dimension implies a negative correlation between response time and error rate, derived by a median split: 75% of the subjects supposedly fall into one of these two categories:

impulsive or reflective; and the remaining one quarter into one of the categories: fast-accurate or slow-inaccurate. Kagan et al have been heavily criticised for eliminating all subjects outside the reflectivity-impulsivity dimension (Block, Block and Harrington, 1974).

This negative correlation between response time and error rate is thought to be a direct measure of delay before responding. However, it is not clear that the reflective subject is actually evaluating the choices more carefully than impulsive subjects (Messer, 1976). Research on scanning behaviour suggests reflectives might be weighing up hypotheses before choosing. For example, Siegelman (1969) studied glancing behaviour in reflectives and impulsives, where subjects were required to press a button in order to focus in on the alternative stimuli. Reflectives were found to look more often and for longer periods at all figures than impulsives.

The MFFT also correlates moderately with the EFT, particularly the error scores: reflectives are more field-independent and impulsives more field-dependent (see Campbell and Douglas, 1972; Neimark, 1975).

In relation to the domain of problem solving, reflectives are superior in performance to impulsives. The problems which have been considered have all introduced an element of uncertainty. Reflectives tend to have better short-term auditory memory on a serial learning task (Kagan, 1966) as well as superior short-term visual memory on a visual recognition task (Siegal, Kirasic and Kilburg, 1973). Ault (1973) and Denney (1973) using variations of Mosher and Hornsby's (1966) 20 question games found that reflectives were more likely to ask questions that quickly eliminated a greater number of possible answers. Consistent with this, Nuessle (1972) found that reflectives made more efficient use of feedback in problem solving tasks. On the whole, reflectives appear to show superior performance on a number of memory and problem-solving tasks.

It has been suggested that anxiety over failure plays a role in performance. For example, Reali and Hall (1970) found that both reflectives and impulsives increased their response time after making an error on MFFT. The performance curve has often been observed in other psychological tests such as the Sternberg memory test and may not be particularly related to the reflectivity-impulsivity dimension.

No work has yet been conducted on the details of the relationship between reflectivity-impulsivity and dialogue style but the findings that EFT scores and dialogue style correlate, and EFT and MFFT correlate moderately, suggest that the reflectivity-impulsivity dimension may be a useful source of information for the interface designer.

2.2.1.3 Serialist-holist

The serialist-holist dichotomy was first coined by Pask and associates in a set of studies which looked at the way students learned a set of facts. The early work involved a set of information cards which students were required to work out the taxonomies of a species of fictitious Martian animals: the Clobbits. While deducing the taxonomies the students gave reasons for their decisions and what they had learned from turning the cards over (TEACHBACK). One set of the students was classified as serialists: they worked in a step-by-step manner, concentrating on simple hypotheses related to just one concept, while holists formed complex hypotheses involving several concepts. Serialists were more narrow in their approach and did not make use of analogies while holists tended to overgeneralise and thus miss important details. The students were then given a new taxonomy to learn (the Gandlemuller) the presentation of which might either match or mismatch the students' learning strategy. Those students in the matched conditions demonstrated a greater understanding of the subject learned. From the serialist-holist dimension two learning strategies were postulated: "The term comprehension learning is used for that facet of the learning process concerned with building descriptions of what may be known. Operation learning is the corresponding term for the facet of the learning process concerned with mastering procedural details" (Pask et al, 1977, p.68).

In these early studies, learning strategies were inferred from the learning tasks but the tasks are in fact inexact and time-consuming (Entwistle, 1978). The Spy Ring History Test attempts to tap three components of learning: comprehension, operation, and versatility. The subjects learn the history of a spy ring and all the intricate relationships within the domain but the task is structured in such a way that these components can be measured. Those students who are classified as "comprehension learners" tend to learn the similarities and differences between a set of networks. "Operation learners" focus on the roles played by particular spies, and "versatile learners" demonstrate a combination of these two types. The latter are able to make better inferences about what will happen next. In terms of learning to use computers, the versatile learner would be predicted to be most adept and in general to be the most successful at interacting with them. The comprehension style of learning is predicted to help users to understand the general overview of how a system works by employing analogies from other domains. On the other hand, the operational style of learning is relevant at the interaction level. These users are able to turn description into actions, that is, they are able to understand how to accomplish a task.

2.2.2 Personality Factors

Van der Veer et al (1985) have suggested some personality factors which might give designers an indication of what characteristics of the interface should be adaptive. Since personality factors tend to remain constant, they can be specified before the interaction occurs.

2.2.2.1 Introversion-extraversion

Three different hypotheses have been proposed concerning the relationship between introversion-extraversion and arousal. There is as yet no experimental evidence which can discriminate between the hypotheses. The hypotheses are:

i. Introverts are more aroused than extraverts but have the same optimal level of arousal (Eysenck, 1967).

ii. Introverts and extraverts have identical resting levels of arousal but the consequences of stimulation differ (Eysenck, 1971).

iii. The arousal levels of introvert and extravert are the same but introverts have a lower optimal level than extraverts.

In terms of research into verbal learning, extraverts have been found to "learn" faster than introverts in tasks which are difficult or which involve response competition but this difference in learning is reduced when the task is simple (Allsopp and Eysenck, 1974; Bone, 1971; Jensen, 1964). There is also modest support for the hypothesis that the period of consolidation is longer for introverts than extraverts thus the short-term retention of extraverts exceeds that of introverts but introverts have better long-term retention (McClean, 1968; Osborne, 1972).

Extraverts may also be less cautious than introverts. Cameron and Myers (1966) found that extraverts show a greater readiness to make a response despite being uncertain as to whether the response is correct or not.

High arousal has been related to divergent thinking and is thought to have an inhibitory affect (Martindale and Greenough, 1973). High divergent thinkers have shown greater incidental learning and may be less aroused than low divergent thinkers.

Van der Veer et al (1985) suggest that extraverts tend to get easily bored with routine tasks and prefer diversity. The introversion-extraversion scale could be used to establish the amount of help to be given and the type of dialogue. One would expect introverts to prefer a computer-initiated dialogue with plenty of feedback as to what is going on in an interaction in order to maintain a low level of anxiety. Extraverts would soon become bored with this type of interaction and would want to diversify at an early stage in the interaction. Extraverts would wish to make an earlier transition from a computer-initiated language to a user-initiated language compared to introverts. On the other hand, in studies investigating the relationship between personality types and academic progress, introverts have consistently been found to be more successful than extraverts. Extraverts high in motivation and study methods can do comparatively well (Entwistle and Wilson, 1977).

Hoe, Poupeye and Vandierendonck (1987) in a study investigating the optimal breadth and depth of menu-driven systems found that extraverts were significantly more accurate than introverts, particularly during the first part of the trials. They account for these differences by suggesting that introverts tend to avoid new situations while extraverts do not. In addition, they found a similar difference between high and low scorers on the neuroticism scale. The authors, however, do not specify whether the scales were examined independently or not. For example, does a neurotic extravert perform in the same way as a stable introvert?

2.2.2.2 Anxiety

The findings of research on anxiety and memory have been less conclusive than those of the introversion-extraversion dimension. One problem concerns the way in which anxiety is assessed, that is, by administering questionnaires which assume that anxiety is unidimensional (Eysenck, 1977). The theory now distinguishes between <u>state</u> anxiety which is transitory and <u>trait</u> anxiety which is a stable characteristic of a person. Under stressful conditions high trait anxiety subjects have been found to respond with more state anxiety than low trait anxiety subjects (Hodges and Spielberger, 1969; Rappaport and Katkin, 1972) but only if the stress constitutes a "psychological" threat to the person's self-esteem (Hodges and Spielberger, 1966; Katkin, 1965).

In research on anxiety and verbal learning, paired associates tasks have been employed. When no strong competing responses exist, the high anxiety group shows superior learning over the low anxiety group (Spence, 1958; Taylor, 1958). This finding is not always supported (Kamin and Fedorchak, 1957; Lovaas, 1960). On the other hand, when competitive paired associates lists are used (some of the stimulus items are initially more closely associated with response terms other than those with which they are paired with in the original list) the high anxiety group tends to do poorly at the beginning (Ramond, 1953; Standish and Champion, 1960) but may surpass the low anxiety group as learning progresses (Standish and Champion, 1960).

One problem with a dimension such as "anxiety" is that the majority of users will fall somewhere in between the two extremes and the dimension may only be of importance to human computer interaction in a minority of extreme cases.

Anxiety, negative attitudes and degree of motivation affect performance. A user's cognitive and learning style is also a good predictor of performance, but from a different standpoint. Cognitive and learning styles look at the most natural way an individual user will approach a particular problem and therefore the emphasis is on the user's problem-solving abilities. On the other hand, personality factors work in the opposite direction, emphasising the structure of the environment within which the user has to interact such as how supportive the dialogue structure is. This suggests that strong correlations may exist between cognitive and learning styles and personality factors, and dialogue styles. The degree of categorisation of types of users will suggest dimensions on which to base adaption in the computer interface.

Study 1 (Chapter 3) is an exploratory study investigating the effects of one cognitive styles test (the Embedded Figures Test) on performance on a computer task. The usefulness of cognitive and learning styles testing and personality inventories are further explored as measures of a user's success in the domain of human computer interaction in Studies 4 and 5.

Chapter 3 The Relationship between Field Dependence and Performance on a Computer Task with Varied Amounts of Support

3.1 Introduction

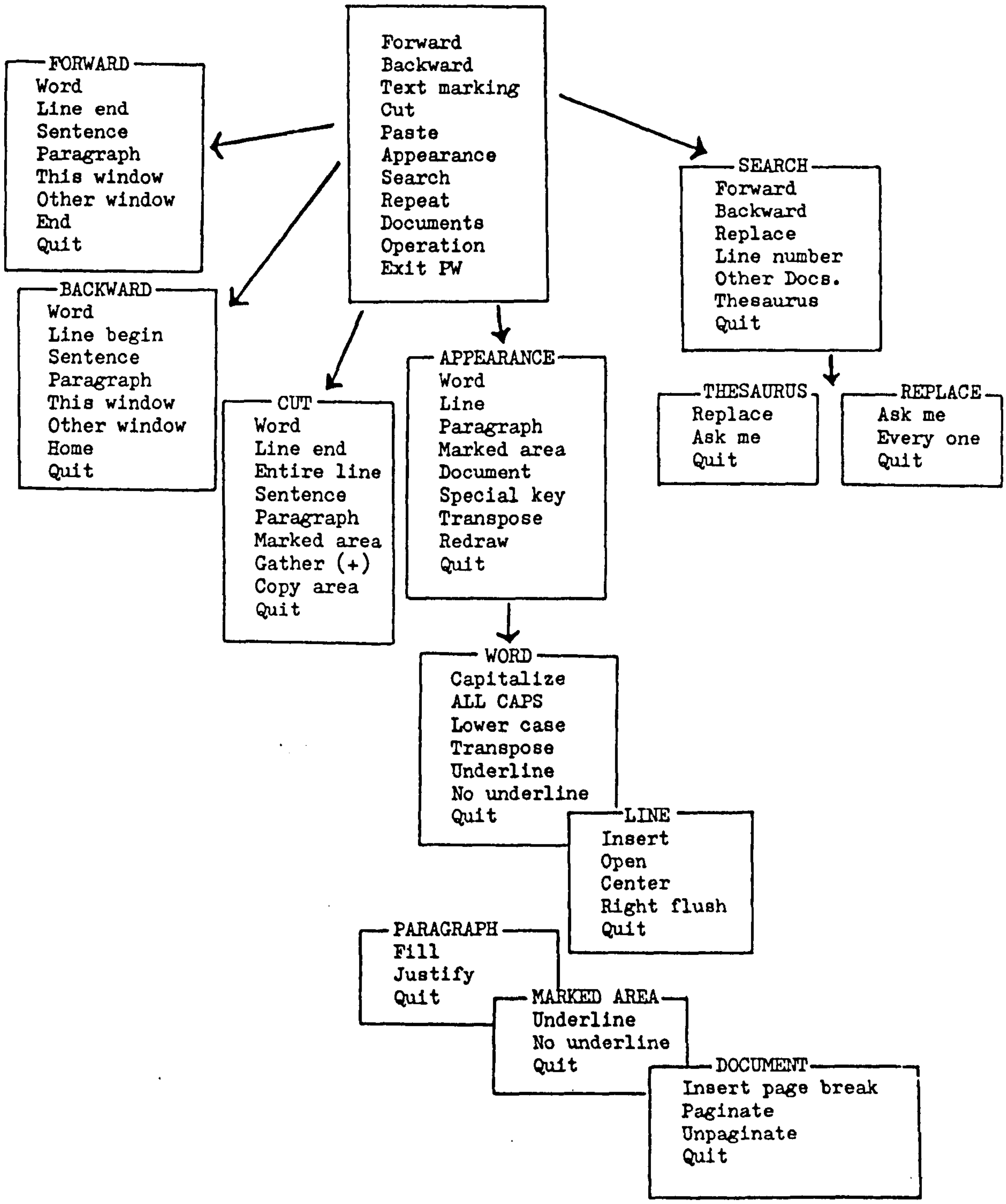
Chapter 2 has highlighted the reasons surrounding the desirability of adaptive interfaces and how they can be achieved from a user-centred viewpoint. However, few attempts have been made to investigate how cognitive style can affect performance in human computer interaction. Fowler et al (1986) found differential effects of field dependence-independence as measured by Witkin's EFT on two types of interactive languages which differed in the amount of support they provided (see Chapter 2 for more details). Performance was measured in terms of errors and number of moves. However, the amount of help requested can be measured independently of the environment which the language provides and thus may also be important in differentiating between users. In particular one would expect users who are field dependent to require more help to complete their task. In the study reported below, amount of help requested is measured, in addition, to time on task and number of moves.

Many aspects of human computer interaction involve a good deal of problem solving. This study uses a map reading task which can be viewed as similar to the types of problems encountered when working one's way around a word processor. For example, many word processors use embedded menus which involve the user making choices as to how to proceed. Often the user receives very little state information or feedback as to where s/he is in a hierarchy of menus (see Sheehy, Forrest and Chapman, 1988). This problem has frequently been documented (cf. Nievergelt, 1982). Lack of feedback from the system leaves users wondering "where am I?". This problem is particularly common in systems which utilise layers or levels of menus. Colour coding of particular parts of the interface appears to offer a potentially useful solution because it can provide information about the current level of menu and thereby indicate a user's position in an non-intrusive way. Study 2 examines more closely the potential of employing types of coding to augment performance.

Figure 3.1 is a schematic diagram taken from the Perfect Writer II word processing manual. Users have to weave their way around the hierarchy of menus in order to select the appropriate commands. Although the menus are structured hierarchically (i.e., their propositional representation from a designer's perspective), it is doubtful whether users represent the structure in this way. They experience it procedurally and are likely to represent it as such, that is, as procedures of moves within a complex structural environment.

Rogers, Leiser and Carr (1988) have explored the notion of a "building metaphor" as an extension of the "desktop metaphor". They found that users of this system understood more about the operations in terms of the building than

Figure 3.1 Perfect Writer's Menu Tree



they did about the operations of the desktop metaphor or system metaphor.

The idea of a building metaphor is that users of the system are able to move throughout the building in much the same way as they do in every day life. The building metaphor encompasses a greater area than the desktop metaphor and has more scope for different activities. In addition, it also possesses a hierarchical structure. In order to simplify the areas of the building metaphor it is possible to "cue" the users, for example, using colour to provide feedback as to the user's location in the building.

The study described below combines all these ideas. It uses a simplistic version of the building metaphor and colour to provide feedback to the user. It also explores the idea that there are different types of users and that the level of support they require during a human computer interaction can be differentiated by their EFT scores.

3.2 Aims

The aims of this study are:

i) to investigate the relationship between a map reading task and the degree of field dependence that person shows as measured by the EFT.

ii) to investigate the potential application of employing colour as a form of support to the subject to indicate where s/he is in the building.

iii) to predict subjects' performance in the two conditions (colour and no colour) by their EFT scores.

iv) to see if any difference in performance is short-lived.

From the literature on the EFT which demonstrate the differential performances of field dependents and independents, it is predicted that: (1) EFT scores will reflect a subjects' performances on the map reading task. (2) Field independent subjects will perform equally well irrespective of whether they receive support or not. (3) Field dependent subjects perform better when support is provided than when no support is provided. (4) Both field independent and field dependent subjects do equally well if support is provided. (5) Field dependents may not perform well when no support is provided in particular at the onstart, but they begin to catch up with the field independents at a later stage. (6) Since females have a tendency to be more field dependent than males then females are not expected to perform as well as males (cf. Fowler and Murray, 1988, for a fuller account).

3.3 Methodology

3.3.1 Subjects

28 subjects (14 males and 14 females) participated in 2 sessions on 2 consecutive days. Sessions lasted between 15 minutes and 45 minutes. Subjects comprised undergraduate and postgraduate students, research and technical staff of the Department of Psychology at the University of Leeds.

3.3.2 Apparatus

The map reading task program was written by the author in Microtext and run on an IBM PC XT computer (see Appendix 3.1). The program consisted of frames which inform the subject of his/her location; for example, the name of the corridor or intercept or department. It also permits the subject to make a choice as to how to proceed. For example, turn left, turn right, or continue etc. (see Appendix 3.2 for examples).

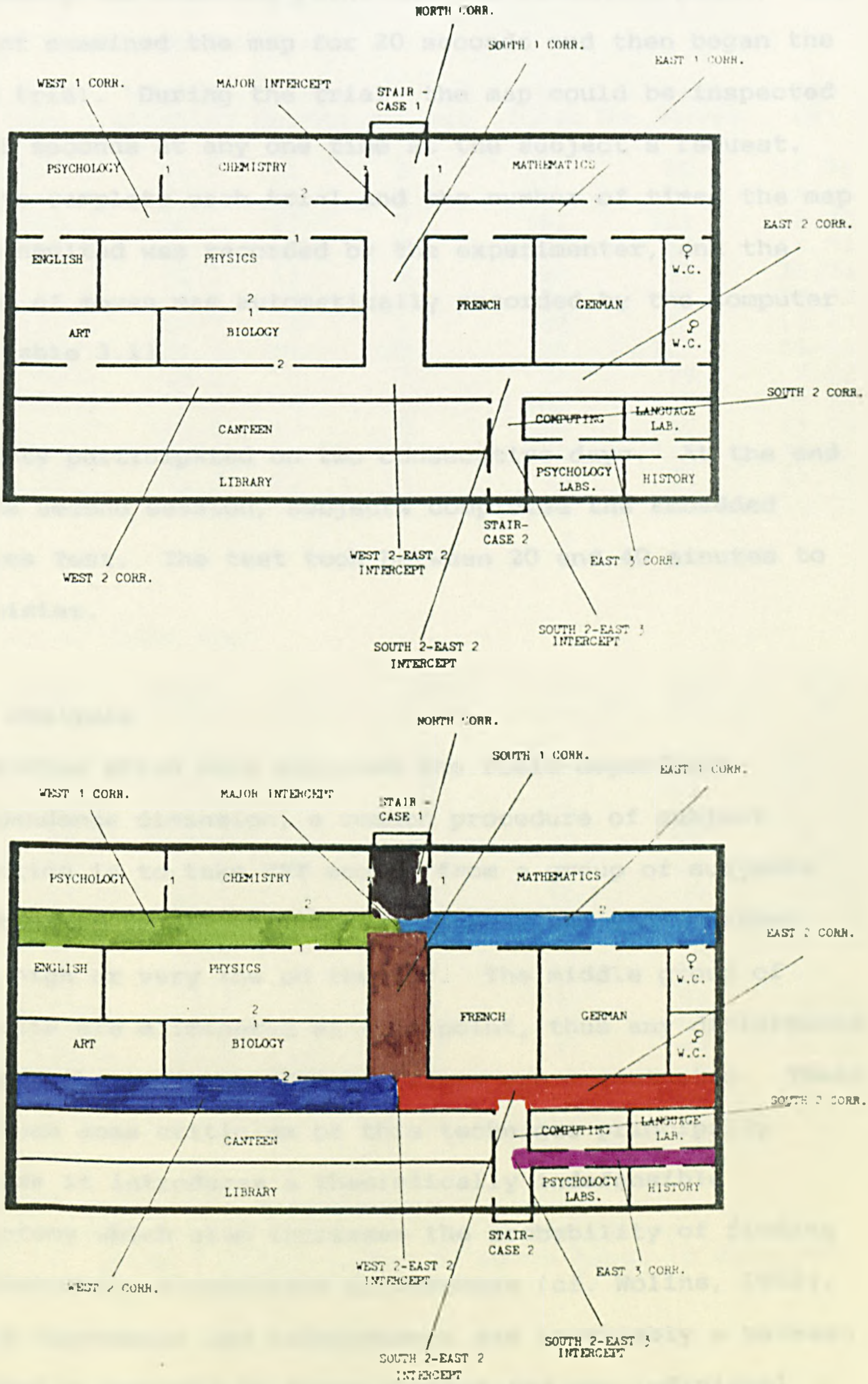
3.3.3 Procedure

The design was between subjects, with equal numbers of males and females assigned to each of the two conditions: colour or no colour. In the colour condition, the corridors were coloured, with the intercepts showing a combination of the colours of their respective corridor. Performance measures in terms of number of moves made (i.e., choice of direction), number of times map was consulted, time to complete each trial were taken.

Subjects were instructed to find the shortest route between two points on the map by using the smallest number of moves and to consult the map as little as possible. Subjects in the colour condition were, in addition, instructed to use the colour to help them know where they were on the map (Figure 3.2).

Subjects were seated in front of the computer. Before each of the five trials, the subject read the text on the screen

Figure 3.2 Colour-coded and No Colour-coded Maps



describing the starting point and the finishing point. The subject examined the map for 20 seconds and then began the first trial. During the trial, the map could be inspected for 15 seconds at any one time at the subject's request. Time to complete each trial and the number of times the map was consulted was recorded by the experimenter, and the number of moves was automatically recorded by the computer (see Table 3.1)

Subjects participated on two consecutive days. At the end of the second session, subjects completed the Embedded Figures Test. The test took between 20 and 40 minutes to administer.

3.4 Analysis

In studies which have explored the field dependence-independence dimension, a common procedure of subject selection is to take EFT scores from a group of subjects and then to select only those subjects who score either very high or very low on the EFT. The middle group of subjects are eliminated at this point, thus any differences between the two experimental groups are accentuated. There has been some criticism of this technique principally because it introduces a theoretically indefensible dichotomy which also increases the probability of finding statistically significant differences (cf. Wolins, 1982). Field dependence and independence are invariably a between subject's variable in these studies and any individual differences are ignored. If significant differences

Table 3.1 Subjects Performance on Computer Task

Colour Coded Maps Condition

	Day1 Time	Day2 p/trial($\bar{x}$)	Day1 No.Map	Day2 request	Day1 Total No.	Day2 Moves	EFT ($\bar{x}$)
S1(m)	137.1	81.7	11	6	107	88	14.96
S2(m)	65.9	55.5	6	5	71	81	14.15
S3(m)	186.1	90.5	10	7	161	121	36.40
S4(f)	209.5	87.1	10	2	139	63	15.32
S5(m)	132.5	58.7	11	2	89	67	25.74
S6(f)	166.2	68.8	7	4	154	84	27.02
S7(m)	121.5	42.7	3	1	103	71	10.32
S8(f)	178.7	229.9	32	23	169	201	59.55
S9(m)	62.6	28.0	0	0	73	59	8.93
S10(f)	100.9	70.4	8	5	100	106	9.98
S11(f)	132.2	103.8	16	8	108	117	15.14
S12(f)	101.0	80.0	4	3	123	116	22.12
S13(m)	53.9	42.4	22	11	59	63	42.78
S14(f)	111.6	79.2	10	4	139	103	24.75

Non-colour Coded Maps

	Day1 Time	Day2 p/trial($\bar{x}$)	Day1 No.Map	Day2 request	Day1 Total No.	Day2 Moves	EFT ($\bar{x}$)
S1(m)	86.7	63.9	5	4	73	73	20.17
S2(f)	61.9	41.0	12	8	77	57	33.66
S3(f)	100.7	40.8	13	3	89	63	14.34
S4(m)	58.6	32.1	6	0	75	57	14.61
S5(m)	50.7	50.9	0	0	55	64	11.08
S6(m)	87.4	64.9	7	2	65	59	13.56
S7(f)	194.7	172.0	8	6	120	132	22.77
S8(f)	75.6	40.4	8	3	72	59	10.86
S9(m)	87.9	50.9	5	1	83	71	19.9
S10(m)	112.1	65.3	5	4	99	83	10.65
S11(f)	118.8	40.8	9	5	116	63	32.96
S12(f)	123.8	51.5	14	2	135	67	21.85
S13(m)	97.0	41.3	6	0	73	61	9.85
S14(f)	86.1	39.4	5	0	71	59	14.62

Key:
Time p/trial = mean time per trial
No. Map request = frequency of map consultation
Total No. Moves = total number of moves taken to complete task

cannot be observed over the entire continuum then one can conclude that the differences are too small to have any effect as a tool for diagnosis and is therefore more powerful than employing two dichotomous groups. The use of the raw EFT scores may provide a more accurate picture particularly since it is more in keeping with the idea that field dependence-independence represents opposite ends of a continuum.

The data in this study were analysed using two statistical models. The first uses field dependence-independence as a between subjects variable and the second as the dependent variable. The results are contrasted to demonstrate how different criteria affect the findings and hence one's conclusions.

3.5 Results

A four-way analysis of variance was conducted with sex, style and support as between subjects variables and day as the within subjects variable. 2-tailed tests of significance were used throughout.

Mean time: There were no differences in mean time between subjects who received the colour condition and those who received the no colour condition, $F(1,20)=2.81$. There was no effect of style $F(1,20)=2.45$ or of sex on mean time, $F(1,20)=2.38$. There was an effect of day, $F(1,20)=28.84$, $p<0.0001$, demonstrating that subjects performed significantly better on day 2 (almost all subjects improved on day 2). There were no interactions of day with sex, support or style. It appears that subjects improve irrespective of sex, support and style.

Request for map: There were no differences in frequency of map consultation between groups in terms of sex $F(1,20)=1.81$, style $F(1,20)=1.60$ or support $F(1,20)=2.64$. There was an effect of day, $F(1,20)=53.48$, $p<0.0001$ demonstrating that the frequency of map consultation decreased on day 2 for most subjects.

Number of moves: The between subjects variables, sex and support had significant effects on the number of moves that subjects made $F(1,20)=4.88$, $p<0.05$ and $F(1,20)=7.61$, $p<0.02$ respectively, while style did not, $F(1,20)=3.13$. Males tended to take a fewer number of moves to complete the trials than females. Subjects in the no colour condition tended to out-perform the subjects in the colour condition, suggesting that the colour may have had an adverse effect on performance. Alternatively, there may have been some individual differences masking an effect which could be detected by a repeated measures design. Day also had a significant effect on the number of moves, $F(1,20)=7.84$, $p<0.02$ but again there were no interactions with sex, style or support.

The lack of a significant difference between field dependents and field independents on the performance measures led to further analyses with the omission of the variable sex as this would increase the number of values in each cell (group).

Mean time: There was a significant effect of style on mean time to complete the trials $F(1,20)=4.64$, $p<0.05$. As expected, field independents took less time than field dependents.

Request for map: There were no between subjects effect of support, $F(1,24)=2.82$ or style $F(1,24)=3.36$.

Number of moves: There were significant effects of support $F(1,24)=6.46$, $p<0.02$ and style, $F(1,24)=5.72$, $p<0.05$ on total number of moves. Subjects in the no colour condition performed better than subjects in the colour condition and field independents did better than field dependents. Figures 3.3-3.5 show the effect of style on the dependent variables.

The main effect of sex has been to reduce the effects of style on the performance measures. As a group, there are significant differences between field dependents and independents but not when the group is divided into males and females. The fact that sex influences the findings suggests that the female field independents do as well as male field independents and similarly for field dependent females. This is supported by the results of a regression analysis which indicate that EFT scores cannot predict the sex of the subject.

An alternative approach to data analysis is to use the raw

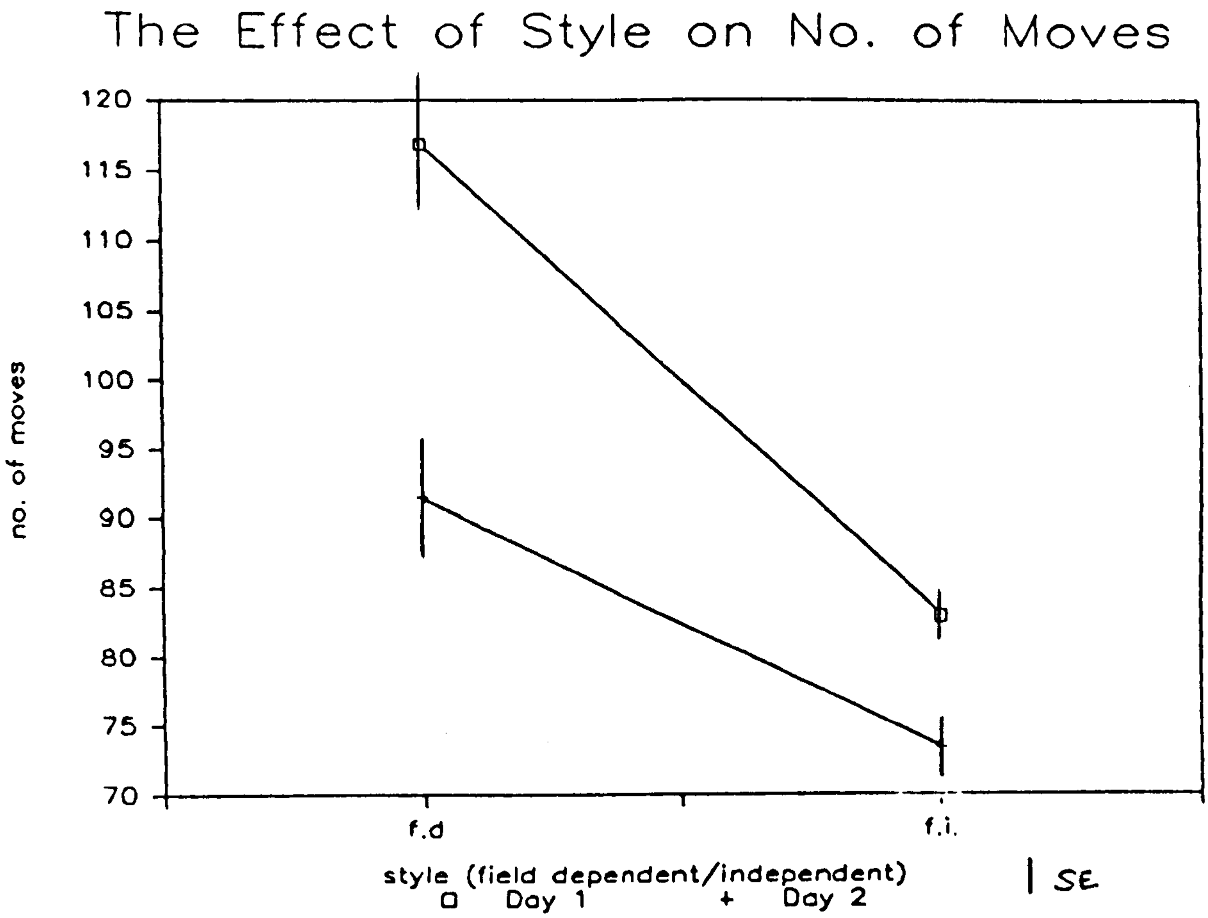

The Effect of Style on No. of Moves
no. of moves
120
115
110
105
100
95
90
85
80
75
70
f.d
f.i.
style (field dependent/independent)
Day 1
Day 2
SE

Fig. 3.3

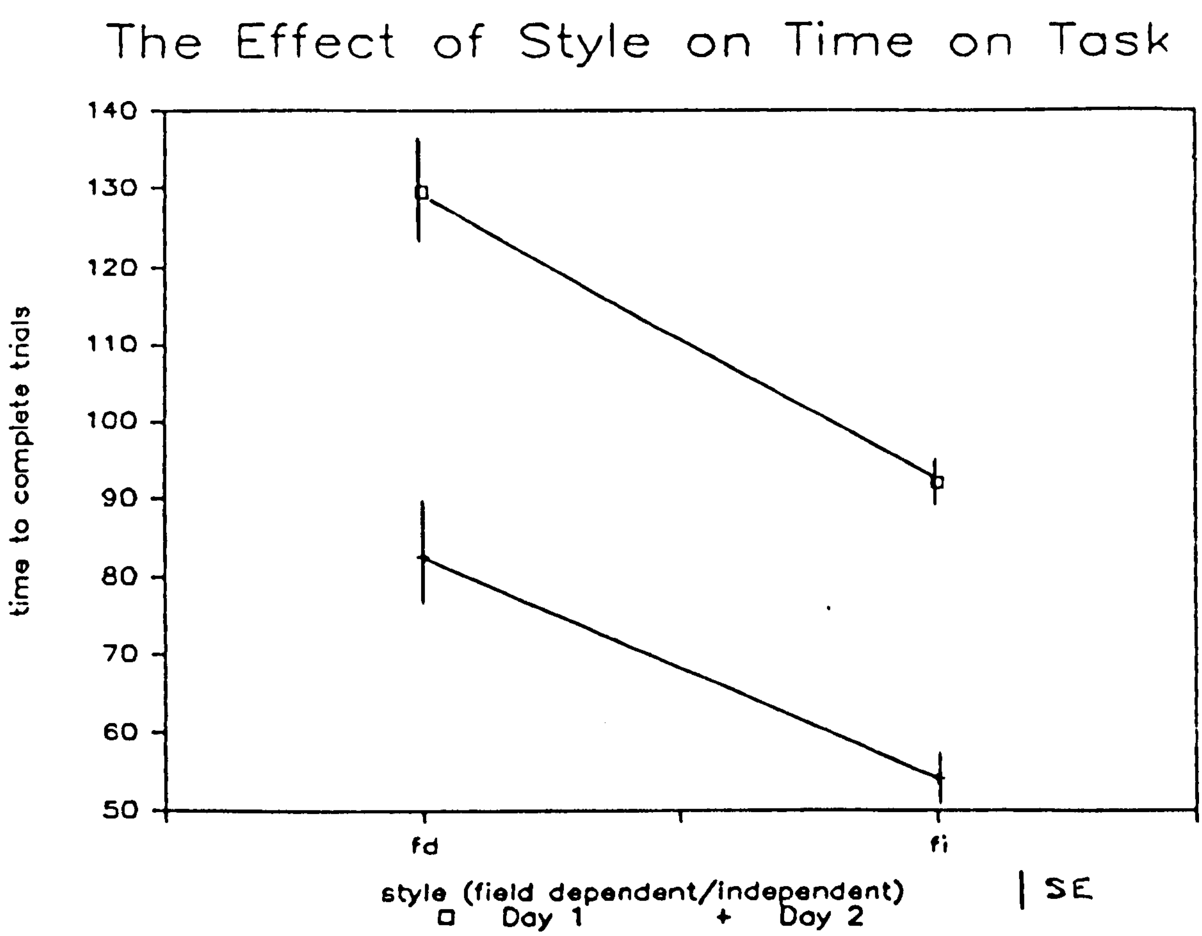

The Effect of Style on Time on Task
time to complete trials
140
130
120
110
100
90
80
70
60
50
fd
fi
style (field dependent/independent)
Day 1
Day 2
SE

Fig. 3.4

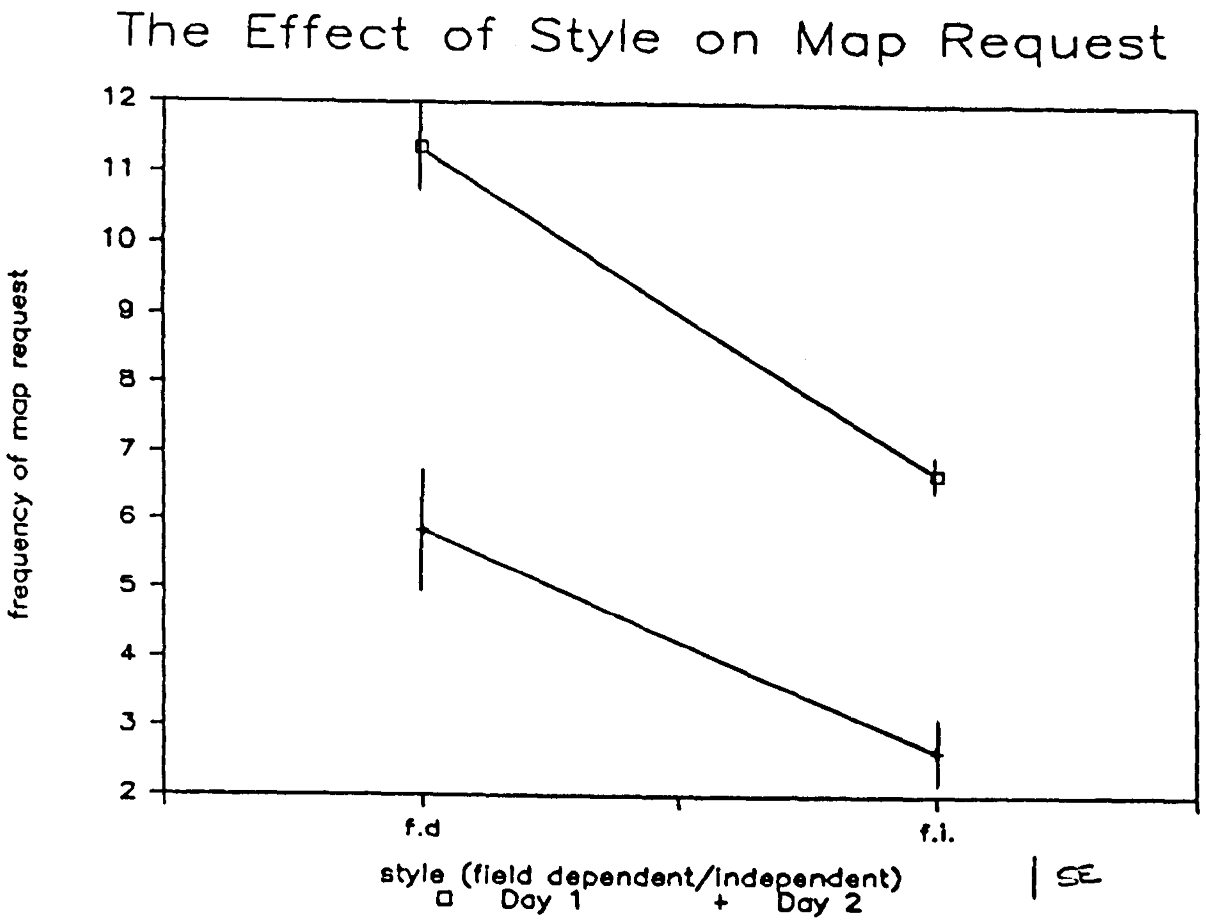

The Effect of Style on Map Request
12
11
10
9
8
7
6
5
4
3
2
frequency of map request
f.d
f.i.
style (field dependent/independent)
□ Day 1 + Day 2
| SE

Fig. 3.5

EFT scores rather than categorising subjects into two groups, as was done here. This may be preferable, especially when subjects have not been selected from the extreme ends of the field dependence-independence dimension. However, if the categorisation of subjects has been satisfactory, then one would expect similar results to emerge when the EFT scores are used.

Two multiple regression analyses were performed on the data for day 1 and day 2. EFT scores were used as the dependent variable, and number of moves, number of times the map is consulted and mean time to complete each trial were used as predictors.

On both day 1 and 2, the number of times the map was consulted was found to be a reliable predictor of EFT scores, $F(1,26)=38.08$, $p<0.00001$ and $F(1,26)=51.62$, $p<0.00001$ respectively. Mean time to complete the task and number of moves did not emerge as significant predictors.

These results suggest that the only factor that can differentiate between field independents and dependents is their request for help. Those subjects who tended to be field dependent consulted the map more frequently.

If one contrasts the findings from the regression analyses with those from the analyses of variance, one finds that different performance measures emerge as being important in differentiating between field independence and dependence.

The difference is due to the "handling" of subjects. The categorisation (in ANOVA) and non-categorisation (in multiple regression analysis) of subjects into two groups have produced contrasting findings. Since subjects in this study were not selected from the two extremes of the field dependence-independence dimension, the categorisation of the subjects tends to be rather arbitrary since the split of the subjects is dependent on the number of participants. Thus, using the raw scores appears to be preferable.

Combining the results from the analyses, the findings suggest that: (1) in general, field independents do better than field dependents. Field independents tend to take less time, fewer moves and require less support. (2) Males tend to perform better than females but not exclusively.

(3) The application of colour does not help either the field dependents or independents. Performance was always superior in the no colour condition. (4) Most subjects improved on day 2 independently of sex, style and support. Those who did not improve tended to be either field independents who were already performing at an optimal level or field dependents.

3.6 Discussion

Research into cognitive styles tests and performance in relation to computer systems is sparse, in particular, when investigating the amount of explicit help or support required by users differing in their style of learning. The results of this study suggest that EFT can predict performance and the amount of support required. Fowler et al (1986) suggested that field dependent subjects begin to "catch up" with their field independent counterparts on the second day in terms of number of moves. The results of this study do not support their conclusion. The majority of subjects improved substantially on day 2, but field independents maintained their lead. One reason for the differences in finding is that the task used by Fowler et al employed verbal operations while the task employed here relies more heavily on spatial abilities.

The implications for such research is that its application to adaptive interfaces could allow users to use strategies which are consistent with their own cognitive style, thus making systems more personalised. The nature of an individual's cognitive style might be inferred from the amount of support and help they require and used to inform the provision of support-levels at a later time.

An alternative explanation for these results is that short-term memory plays a major role in tasks of this type. The reason that some subjects consulted the map less often than other subjects may be due to their superior short-term memory for spatial and shape information. Findings from studies investigating field dependence and memory have found that field dependents have less efficient memory especially when the memory load is high (Robinson and

Bennink, 1978) or when there is interference (Berger, 1977). In this study, memory load was probably high and this may have contributed to the differences in performance between field independents and dependents, in particular, the frequency of requesting for support (the map).

The type of memory task inherent in this study and also in human computer interaction appears to be associative in nature. In this study, subjects had to learn the routes of the map by associating the colours or names of the corridors with the position or orientation of the corridors. The results of studies which have investigated associative learning have been somewhat contradictory and dependent on the task criteria. Most studies have found no relationship between associative learning in terms of paired associates tasks (e.g., Adcock and Webberley, 1971; Mulgrave, 1965) especially when intelligence has been partialled out. Witkin, Dyke, Faterson, Goodenough and Karp (1962) suggest that the significant relationship between field dependence and IQ scores is mainly due to those components in the tests which require analytic functioning. There is also evidence to suggest that learning differs according to how the information is structured (Fleming, cited in Goodenough, 1976). Other findings suggest that field dependents are poor at providing cues (e.g. Frank, 1983). These findings support the notion that field dependents are poor at restructuring and analysing information (Witkin, Moore, Goodenough and Cox, 1977). The next study explores the notion that field dependents are poor at restructuring information in a paired associates test where subjects must provide a mnemonic to help them remember the room they are in.

Study 1 also explored the notion that colour can be useful as a feedback cue to the user, giving information about a user's position, be it on a map or within a software package. Research into the usefulness of colour coding over achromatic coding has primarily been through search and identification tasks. Its usefulness in search tasks has been demonstrated but not its usefulness in identification (see Chapter 5) tasks.

The results of this study concerning the use of colour are discouraging. They suggest that the colour did not provide additional feedback to users. One reason for this finding may be that users found the colour distracting and confusing because it was overused. It may be possible to use colour in the interface as a non-obtrusive background coding to describe one's position within a hierarchy of menus or displays, or within components of the system rather than many colours on one screen at one time. The difference between this type of coding and the one employed in this study is that colour is used more globally. The use of colour in this study tended to be specific. In this case, colour should prove effective as a means of knowing where one is within a system if the user is made aware of its use. The usefulness of colour coding the interface is explored further in the next study.

Study 2 (Chapter 4) explores the role of memory to aid performance. The study also investigates the usefulness of colour as a support to the user, comparing colour with other forms of coding. If memory is in fact a contributor to performance, then interfaces need to support users by reducing their memory loads still further. This may be achieved by employing colour.

3.7 Conclusions

The EFT appears to be able to predict performance, in particular cognitive style appears to be related to the amount of help that a person requests. Short-term memory or associative learning may also play a role, influencing the frequency with which the subject requested help from the map.

Chapter 4 Field Dependence and Associative learning

4.1 Introduction

The findings of studies investigating the role of field dependence in memory and associative learning tasks have found that field independents are superior to field dependents when the task requires subjects to restructure information. In a study by Fleming (cited in Goodenough, 1976), the effect of hierarchical structure within a word list was examined. Words were either presented in a superordinate to subordinate sequence or vice versa (animal, vertebrate, man or man, vertebrate, animal). Since structure aids recall, the subordinate to superordinate sequence is more difficult because the subject has to provide his/her own structure. Thus it was predicted that field dependents would recall fewer words in this condition but there would be no difference in the superordinate to subordinate sequence. This hypothesis was supported.

Similar results were found in a paired associates study by Frank (1983). Field dependents performed equally as well as field independents if the "primer" was the same as those already learned by the subjects but the performance of field dependents was reduced if the "primer" was changed to words which were related differently to the test word. Field dependents performed as well as their counterparts in a free recall test.

The study described below explores the relationship between associative learning and field dependence-independence using the concept of a "building metaphor" as a suitable human computer interface (see Chapter 3). In addition, it is necessary to investigate how users are able to use feedback that is provided by the system as well as their own cues for remembering.

4.2 Aims

The aims of the study were:

i) to investigate the relationship between associative learning and field dependence. This study is similar to that described in Study 1 in that the map structure is maintained as is the use of colour and no colour. The subjects' task is to remember the names or position of rooms on the map by creating a mnemonic for remembering. It is predicted that when a cue is provided, field dependents and independents will perform similarly but when subjects have to provide their own cue, field independents will show superior performance over field dependents. In addition, since females tend to be more field dependent than males, there will be an effect of sex on performance.

ii) to investigate the effect of coding parts of the interface: that is, to examine the effect of non-qualitative (but non-redundant) colour coding, shape coding or no coding on subjects' performance (see Chapter 5 for a review of research on search and identification tasks using colour and shapes as targets). Study 1 demonstrated that colour hindered the users' response, but this may be due to the over-use of colour on the map.

4.3 Methodology

4.3.1 Subjects

Twenty-four subjects (12 female and 12 male students and staff) participated voluntarily in a study lasting between 30 and 45 minutes. Twenty-two of the subjects (11 male, 11 female) also provided EFT scores.

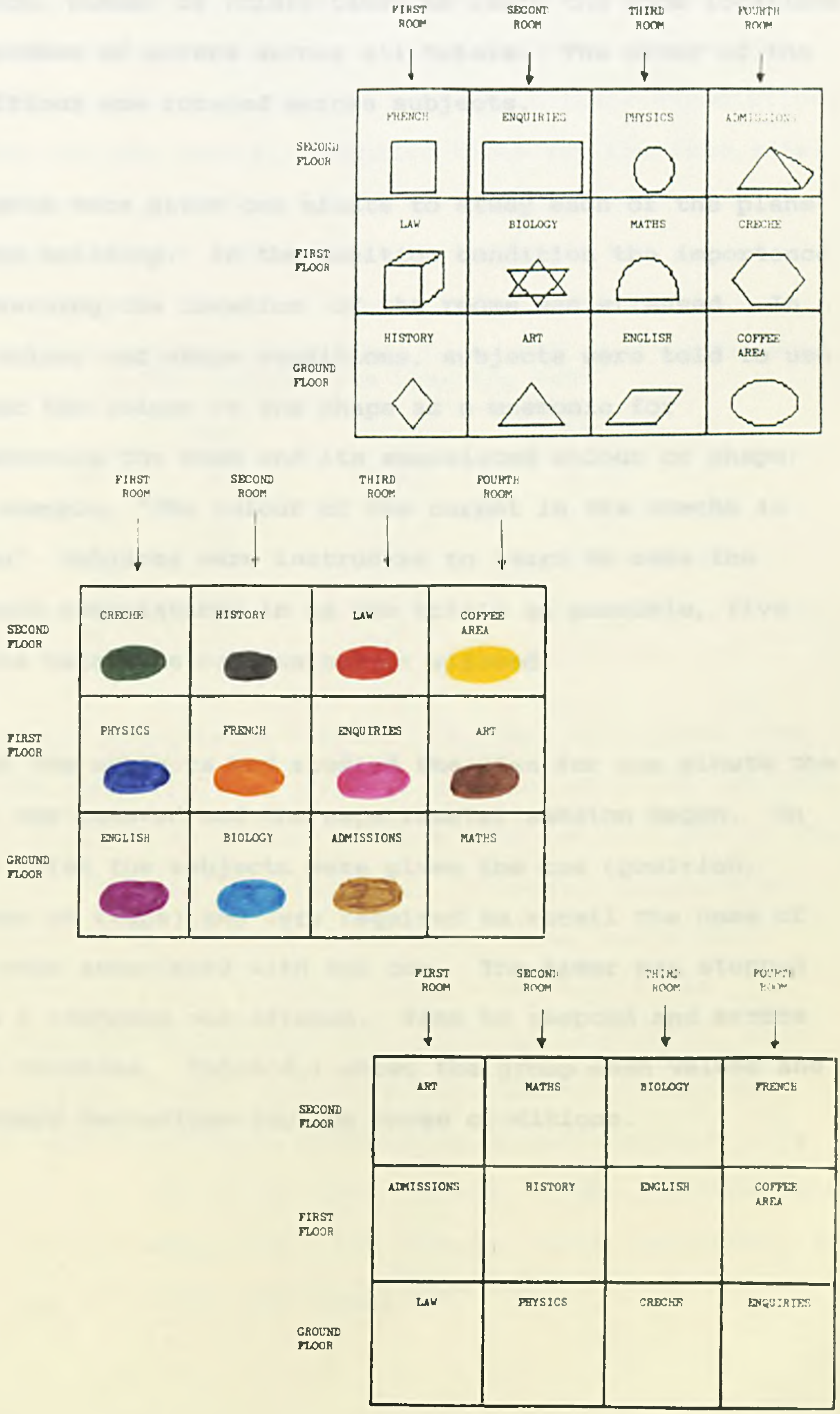
4.3.2 Materials

Three plans of a building were drawn up. The building was in three floors with an equal number of similarly sized rooms on each floor (see Figure 4.1). The plans showed either i) labels of rooms/departments; ii) labels of rooms/departments and associated colours; or iii) labels of rooms/departments and associated shapes. The position of rooms was randomised across the three conditions. Un-labelled rooms were also reproduced on separate cards containing either the floor and room number (position condition), colour only (colour condition) or shape only (shape condition). A timer was used to record the time taken to respond with the correct room title.

4.3.3 Procedure

The study was a repeated measures design with type of cue (position, colour or shape) as a within subjects factor and

Figure 4.1 Plans of Buildings



style and sex as between subject factors. The three dependent variables were: time to respond with correct name of room, number of trials taken to learn the room locations and number of errors across all trials. The order of the conditions was rotated across subjects.

Subjects were given one minute to study each of the plans of the building. In the position condition the importance of learning the location of the rooms was stressed. In the colour and shape conditions, subjects were told to use either the colour or the shape as a mnemonic for remembering the room and its associated colour or shape; for example, "the colour of the carpet in the creche is green". Subjects were instructed to learn to make the correct associations in as few trials as possible, five trials being the maximum number allowed.

After the subjects had studied the plan for one minute the plan was removed and the experimental session begun. On each trial the subjects were given the cue (position, colour or shape) and were required to recall the name of the room associated with the cue. The timer was stopped when a response was offered. Time to respond and errors were recorded. Table 4.1 shows the group mean values and standard deviations for the three conditions.

4.4 Results

The data were examined statistically to see if there were any differences in performance between the 3 conditions irrespective of style or sex. In a majority of cases (87.7%) subjects were able to make the correct associations in five or less trials. Reaction times for the last trial (the learning criterion trial) were used as well as errors and the number of trials taken to learn all the associations.

Table 4.1 Subjects Overall Means and S.D. for Performance on the Associative Learning Task

		Time	Trials	Errors
Colour	$\bar{x}$	1.67	3.33	8.67
	SD	.86	1.31	8.62
Shape	$\bar{x}$	1.56	3.13	7.21
	SD	.83	1.54	7.27
Position	$\bar{x}$	4.47	4.25	15.67
	SD	1.64	1.11	11.47

Reaction times greater than 2.5 standard deviations from the mean were excluded from the analysis to avoid the data being affected by outliers. This resulted in 3.36% of the reaction times being eliminated. A one-way analysis of variance, with repeated measures on three levels was performed on these reaction times. The result was statistically significant $F(2,46)=69.06$, $p<0.0001$. Post-hoc related t-tests were performed, demonstrating that there were significant differences between position only ($\bar{x}=4.47$) and colour ($\bar{x}=1.67$) ($t=9.58$, $df=23$, $p<0.0001$) and position and shapes ($\bar{x}=1.56$) ($t=8.07$, $df=23$, $p<0.0001$), but not between colour and shapes ($t=0.77$).

The same pattern also emerged for the number of errors made $F(2,46)=8.15$, $p<0.001$ and the number of trials taken $F(2,46)=6.04$, $p<0.005$. Post-hoc related t tests showed significant differences between position ($\bar{x}=15.67$) and colour ($\bar{x}=8.67$) ($t=3.36$, $df=23$, $p<0.005$) and position and shapes ($\bar{x}=7.21$) ($t=3.35$, $df=23$, $p<0.005$ but not between shapes and colour ($t=0.7$) for errors made. For the number of trials taken, there were significant differences again between position ($\bar{x}=4.25$) and colour ($\bar{x}=3.33$) ($t=2.74$, $df=23$ $p<0.02$), and position and shapes ($\bar{x}=3.13$) ($t=3.34$, $df=23$, $p<0.005$, but not between shapes and colour ($t=0.58$).

In a two way ANOVA with sex and style as between subjects factors, both sex and style emerged as significant effects on the number of errors made during the sessions $F(1,18)=6.92$, $p<0.02$ and $F(1,18)=5.80$, $p<0.05$. Field independents made fewer errors overall ($\bar{x}=8.66$) than field dependents ($\bar{x}=12.33$). There was no interaction with condition. Inspection of the means suggested that there were no differences in performance in the colour condition (field dependent: $\bar{x}=8.36$, field independent: $\bar{x}=8.55$) but that there were differences between field dependents (fd) and independents (fi) in the shape condition (fd: $\bar{x}=8.73$, fi: $\bar{x}=5.0$) and in particular the position condition (fd: $\bar{x}=19.82$, fi: $\bar{x}=12.45$).

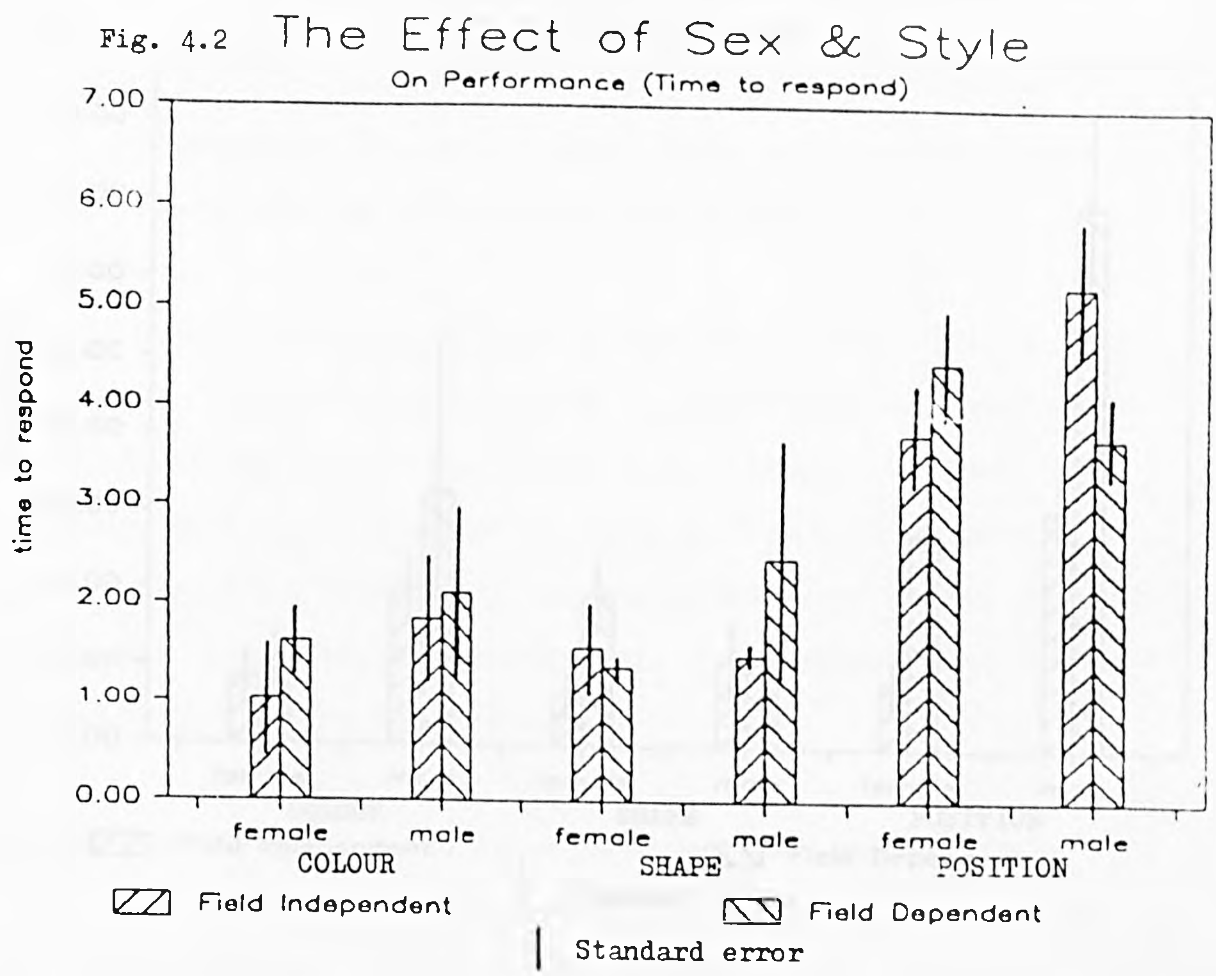
Fig. 4.2

The Effect of Sex & Style
On Performance (Time to respond)
time to respond
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
female
male
female
male
female
male
COLOUR
SHAPE
POSITION
Field Independent
Field Dependent
Standard error

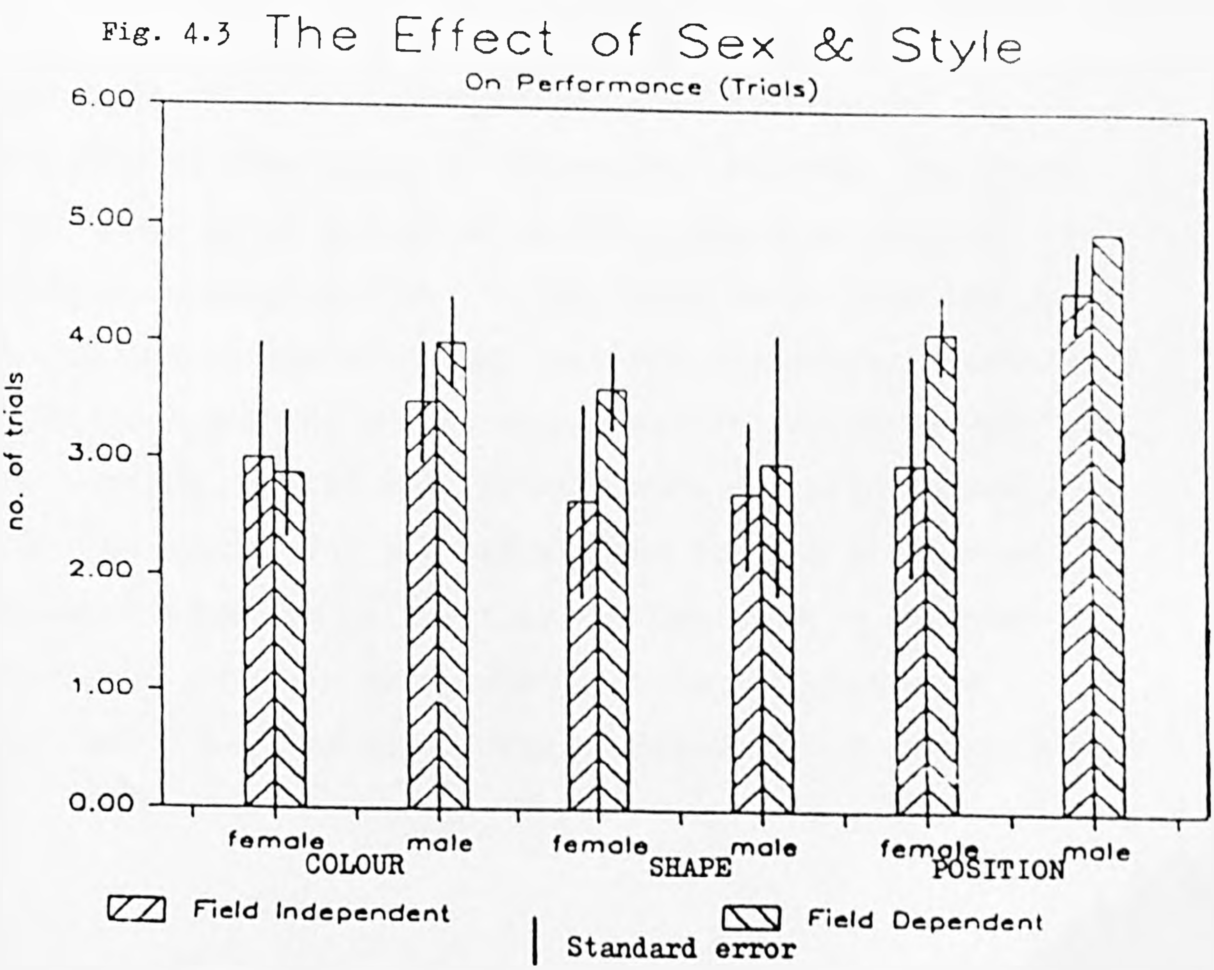
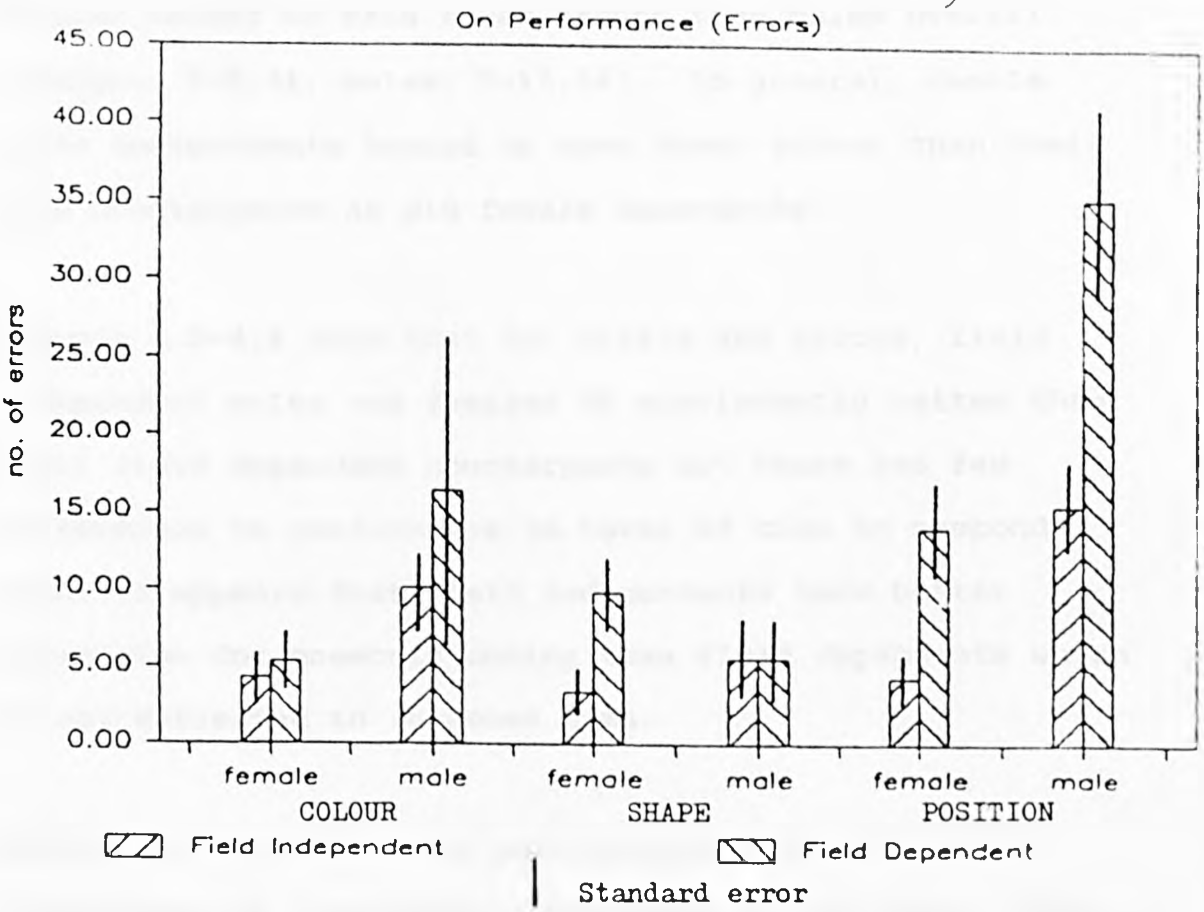
Fig. 4.3

The Effect of Sex & Style
On Performance (Trials)
no. of trials
6.00
5.00
4.00
3.00
2.00
1.00
0.00
female
male
female
male
female
male
COLOUR
SHAPE
POSITION
Field Independent
Field Dependent
Standard error

Fig. 4.4 The Effect of Sex & Style



Females tended to make fewer errors than males overall (females: $\bar{x}$=8.41, males: $\bar{x}$=12.64). In general, female field independents tended to make fewer errors than their male counterparts as did female dependents.

Figures 4.2-4.4 show that for trials and errors, field independent males and females do consistently better than their field dependent counterparts but there are few differences in performance in terms of time to respond. Thus, it appears that field independents have better strategies for mnemonic cueing than field dependents which is not reflected in response time.

Regression analyses were also performed to examine the differences of statistical techniques on the data. Using the EFT scores as the dependent variable, and number of errors, number of trials and time to respond for each condition as sets of predictors, none emerged as significant predictors of EFT/style. However, the fact that there is no effect of sex or style when analysed as separate between factors in the ANOVA shows that the data were somewhat "unusual" and that the significant results from the combined effects of sex and style may be spurious. For example, the Figures 4.5-4.7 show the mean values and standard errors for the two groups: sex and style when analysed separately. Females are observed to perform similarly to field independents while males perform similarly to field dependents in terms of errors and trials

Fig. 4.5 Separate effects of Sex & Style

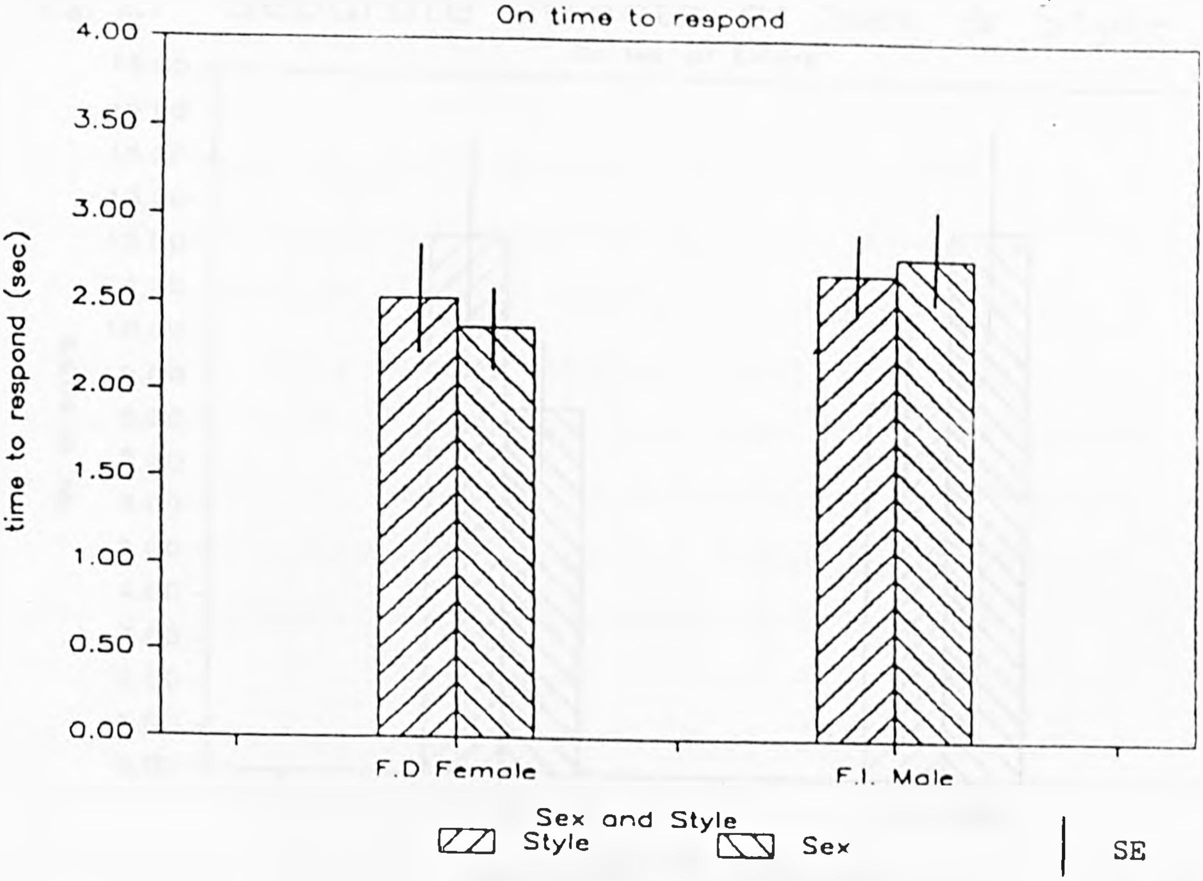
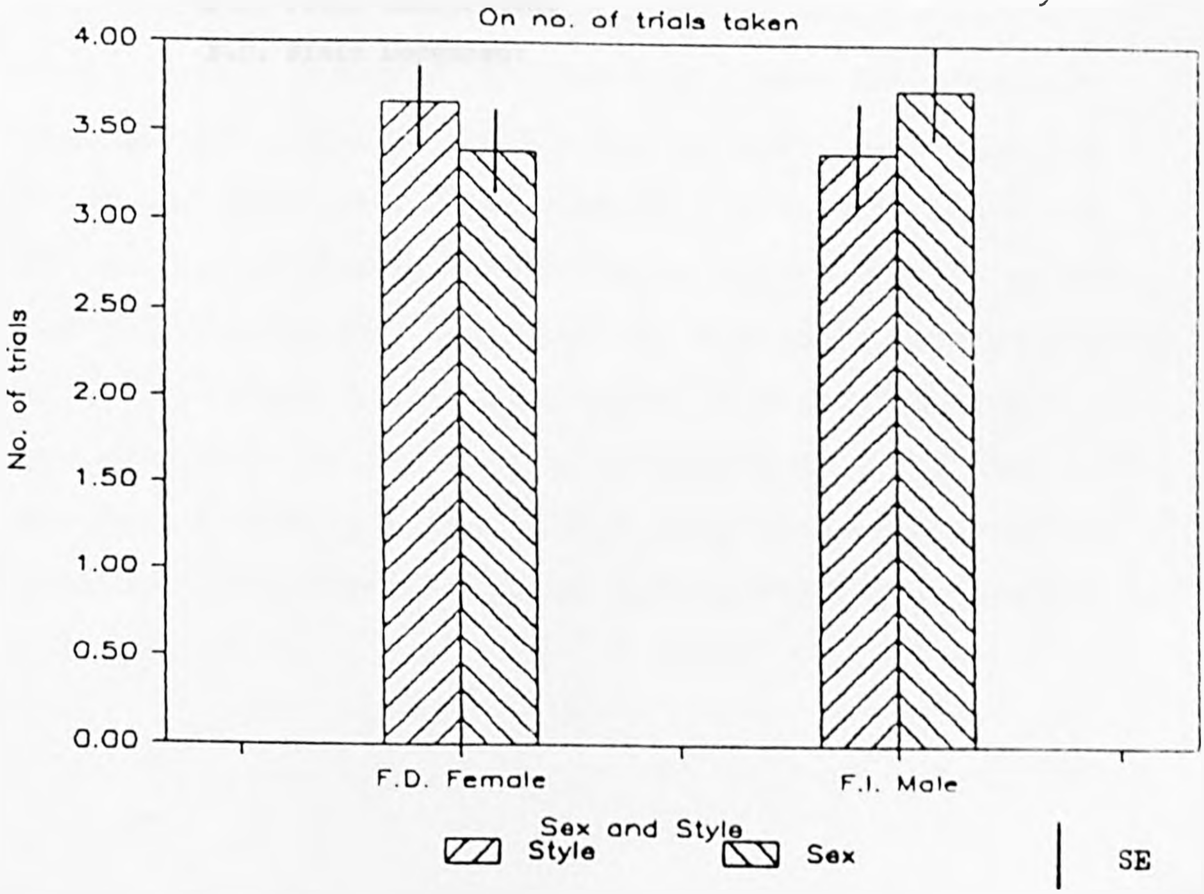


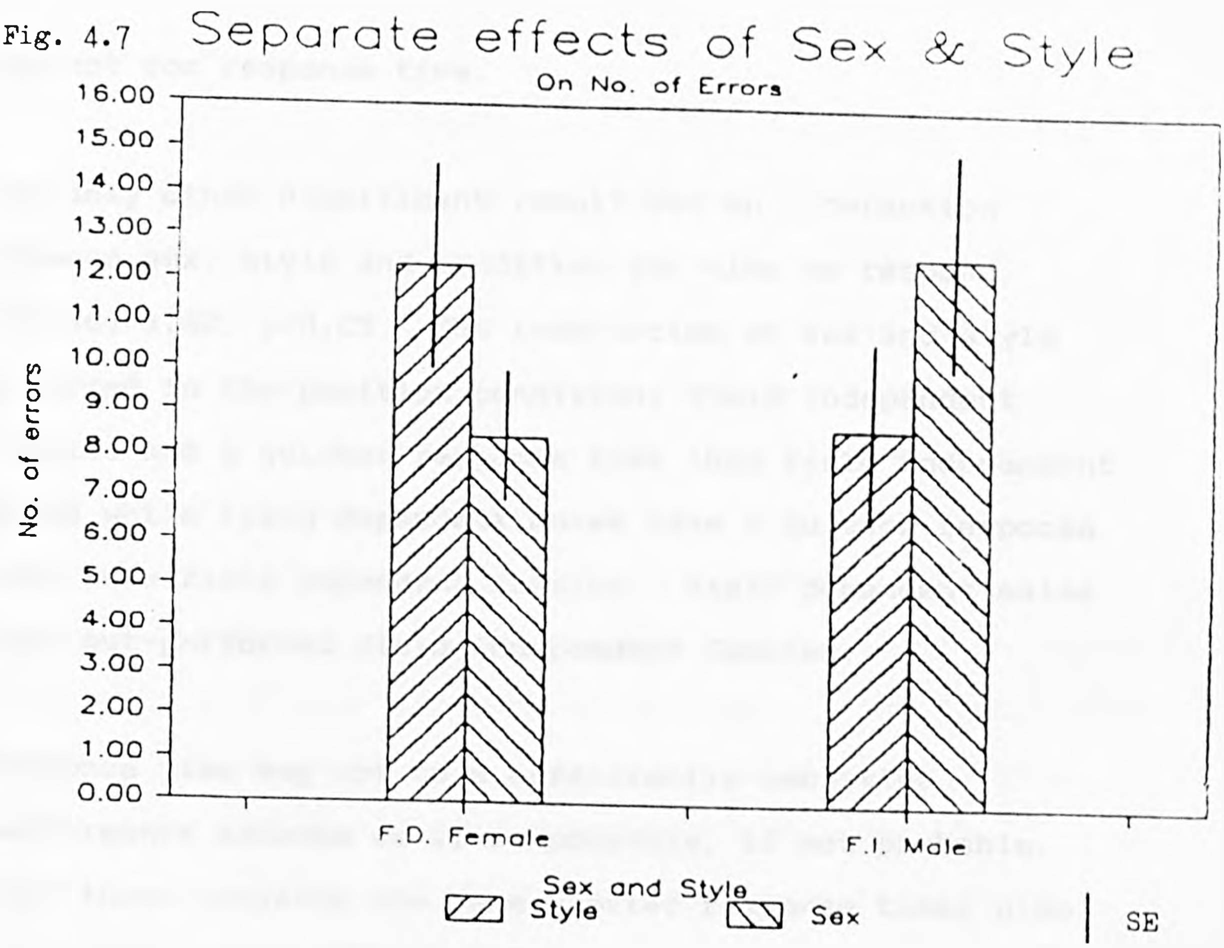
Fig. 4.6 Separate effects of Sex & Style



F.I Field Independent

F.D. Field Dependent

Fig. 4.7



F.I. Field Independent

F.D. Field Dependent

but not for response time.

The only other significant result was an interaction between sex, style and condition for time to respond, $F(2,36)=3.42$, $p<0.05$. The interaction of sex and style occurred in the position condition: field independent females had a quicker response time than field independent males while field dependent males have a quicker response time than field dependent females. Field dependent males also out-performed field independent females.

Response time may not be a sufficiently sensitive performance measure as it is possible, if not probable, that those subjects who have shorter response times also make more errors and may take more trials to learn the associations. The short response times are due to the few that s/he can remember or which might have been learned through rote learning rather than attempting to adopt an effective strategy for remembering. Thus strategies may differ: a performance which shows a high number of errors and trials suggests "guessing" while a low number of errors and trials suggests the employment of a strategy for remembering. This will be particularly true for the position condition which is difficult to learn in a rote fashion. This interpretation is supported by inspection of the data.

4.5 Discussion

In order to draw firm conclusions about the effectiveness of cues for remembering, it is necessary to demonstrate that the three conditions were equally difficult. The position condition differed from the colour and shape conditions in that subjects had to translate the textual information describing the position of the rooms into the corresponding spatial orientation. A strategy or cue for remembering would be useful in this condition. In the colour and shape conditions subjects were required to make associations between the cue and the name of the room. In these conditions a cue was provided. It could be argued that the differences in performance time between the position condition and the colour and shape conditions were due to two different processing strategy used by the subjects. Computer users are often confronted with a similar situation. Usually there are very few cues to allow users to know where they are and they often have to remember system locations by attending to the physical positioning of objects on the screen or by recalling the sequence of displays or paths within the system architecture. On the other hand, users are rarely confronted with objects in isolation, as in the position condition. However, the highly significant result demonstrating a difference between position and colour and shapes remains an important one for it indicates that any type of cue which permits users to make appropriate associations results in better performance. Particular attention needs to be paid to the identification of effective cues.

The absence of performance differences between colour and shapes may have been due to differences in the types of association that can be made between the rooms and their respective colour or shape. Colours are flat and unidimensional and the type of associations that can be made are limited. On the other hand shapes possess more distinctive attributes and it is easier to assign functional uses to shapes than to colours. This suggests that shapes might enhance performance relative to colour cues. This difference was not observed for the reaction time data (colour: $\bar{x}$=1.67, shapes: $\bar{x}$=1.56), or the error data (colour: $\bar{x}$=8.67, shapes: $\bar{x}$=7.2) or the trial data (colour: $\bar{x}$=3.33, shapes: $\bar{x}$=3.13).

The results suggest that visual cues are effective in augmenting performance. An advantage of colour over shape is that colour can be employed in subtle ways which do not distract the user. It is easier, from a computational perspective, to use colour than to use geometric figures and colour may also enhance the overall aesthetics of the display (see Chapter 6, Sheehy and Forrest 1986a) making the screen more comprehensible. The results demonstrate that colour coding of the floors of a building will be useful where such an interface metaphor is used. These results can also be generalised to system design, for example, to the colour coding of screens or menus to help users to make associations which will allow them to know where they are in a system.

The results also suggest that EFT can predict performance (cf. Study 1) and this performance may be related to the ability to produce meaningful cues in order to remember a location. It was suggested that field independents might have better strategies for providing cues in order to remember. Subjects were asked to think of a mnemonic or strategy for remembering. In the colour and shape conditions, a cue is provided, but not a strategy for remembering, while in the position condition, no cue or strategy is provided. It was predicted that field dependents and independents would perform equally well in the shape and colour conditions, but that field independents would do better than dependents in the position condition. Although there was a significant effect of style for the number of errors made, there was no interaction with condition. Field independents made significantly fewer errors than field dependents in all three conditions. An examination of the data shows that general differences do exist for position (fd: $\bar{x}$=19.82, fi: $\bar{x}$=12.45) and shapes (fd: $\bar{x}$=8.73, fi: $\bar{x}$=5.0) but not for colour (fd: $\bar{x}$=8.36, fi: $\bar{x}$=8.55).

Although more males were field independent than females (males=8, females=3), females made fewer errors than males.

This appears contradictory since field independents were found to make fewer errors. This was mainly due to the three field dependent males making a large number of errors ($\bar{x}$=19.12) and the three field independent females making very few errors ($\bar{x}$=4.0). Thus, the sample may have been "unusual" in that the outliers might have affected the overall results. The lack of significant effects of sex and style when the factors are analysed separately and in the regression analyses using the raw EFT scores suggest that the joint effects of style and sex may be due to sampling. A larger sample of subjects is needed to show whether style and sex are important in associative learning tasks.

In sum, the results show that field independent females make fewer errors than field dependent females and that field independent males make fewer errors than field dependent males. Overall, field independents make fewer errors than field dependents and females make fewer errors than males despite being more field dependent than males.

Although there was no interaction of sex and conditions, the raw data suggest that males may be better at using shapes as a cue than females, and females may be better at using colour and creating a mnemonic for remembering position. Thus, there may be a sex difference in the type of information which males and females find easier to process, but this effect may be weak and possibly only discriminable with larger samples.

Field dependence and independence appears to be a quantifiable phenomenon which can be used to understand a user's performance. Sex is also important not only because females tend to be more field dependent than males but because there may be sex-related differences in learning and processing strategies. For example, males may be better at processing shape information while females may be better at processing colour information. Other cognitive styles tests and even personality factors may also help to produce a fuller picture of a user's "style". This is investigated in a naturalistic setting in Study 4.

4.6 Conclusions

The results of Study 2 suggest that field independents make significantly fewer errors than field dependents in an associative learning task, especially when no strategy for remembering is provided. However, these results may be confounded with sex, since females made fewer errors than males. The data show that more females are field dependent than males but the manner in which the data are categorised (median split) and the outliers (good field independent females, poor field dependent males) may have biased the results.

This study also suggests that any visual cue is better than no cue. The previous section discussed how these findings may be applied to computer systems.

The colours employed in this study are redundant and non-qualitative. That is, the colours do not have 'natural' or 'inherent' meanings and, in addition, they do not add any information. Research on the colour coding of displays has suggested that redundant and non-redundant colour can aid search times but there do not appear to be any clear cut results for identification studies. In Study 3, non-redundant, meaningful colour is evaluated with respect to the identification of naturalistic iconic symbols.

Chapter 5 The Use of Qualitative Colour Coding of Meaningful Stimuli

5.1 Introduction

The usefulness of redundant and non-redundant colour coding has yet to be clarified in a satisfactory manner, especially with regard to its application to real world situations such as pertaining to the human computer interface. Research in this area has been diverse: some of the studies used simple search time for target location in relatively artificial tasks while others have utilised more applied domains such as map reading and aircraft control. This makes it difficult to compare studies and to extrapolate between studies conducted in the laboratory and real world situations.

Some of the advantages of colour coding over achromatic coding are:

i) To attract attention to certain parts of the display.

ii) To make a display more legible and comprehensible.

iii) To induce qualitative judgements e.g., red for danger. (see Murch, 1985, Reising and Emerson, 1985).

iv) To make displays more interesting and to increase motivation. Operators of VDUs prefer colour screens but performance is not always related to the colour component (Christ, 1975). In some instances, colour has increased the number of errors (Narborough-Hall, 1985).

Conflicting findings have been the result of differences in task demands. The literature can be divided into two main areas of research: visual search tasks and identification tasks.

5.1.1 Search Tasks

In Christ's (1975, 1984) reviews of studies on colour coding versus other achromatic coding, it was proposed that redundant and non-redundant colour can effectively reduce search times, in particular when unidimensional displays are used. Only alphanumeric coding appears to be superior to colour coding. Colour used in some multidimensional displays may result in interference with other achromatic attributes (Christ, 1984), thus reducing performance on these other forms of coding. Findings are also modified by task demands. For example, search times for a target increases as the number of colours in the display increase (Luria, Neri and Jacobsen, 1986), and when the colours of the background and foreground colours are hard to discriminate (Farmer and Taylor, 1980), but background items which are sufficiently unlike the target colour operate as if no background items were present (Carter, 1982). There is also evidence to suggest that if subjects are not made aware of the colour coding, colour will have no effect (Eriksen, 1952). In general, colour is processed in parallel but there are variations according to task demands. This is illustrated by the fact that set size has no effect (Luder and Barber, 1984) or little effect (17%) on search times (or identification times) for colour displays but increases search times dramatically for monochrome displays (108%) (Carter, 1979).

5.1.2 Identification Tasks

The general consensus is that colour coding does not facilitate identification of objects over achromatic coding (Christ, 1975, 1984; Davidoff, 1987). In fact, shape has been found to be more quickly identified than colour (Zwaga and Duinhower, 1984). These studies have essentially employed non-qualitative colour coding. Colour which can be used to convey inherent meaning might be expected to augment performance.

Ludar and Barber (1984) compared redundant colour coded Cathode Ray Tubes (CRTs) of cockpits of aircraft with monochrome CRTs. Both search and identification tasks were conducted under dual-task conditions. This involved making judgements about the state of components (open, close and emergency) on the CRT while engaged in a tracking task. Results showed that the colour group were faster than the monochrome group on the search task but the faster identification judgements for the monochrome group over the colour group was not significant. On the tracking task, subjects in the colour group demonstrated superior performance to those in the monochrome group. Luder and Barber suggest that this is related to the higher workload experienced by the monochrome group. An alternative explanation would be that the monochrome group found the task less interesting than the colour group (Davidoff, 1987).

Ostergaard and Davidoff (1975) and Davidoff (1986) in an object recognition and naming task used fruit and vegetables in different hues of red and found that colour did not facilitate identification but did when subjects had to name the colour. It is possible that the task created a stroop effect.

Spiker, Rogers and Cincinelli (1986) investigated the use of computer generated topographic displays for army helicopters in place of the conventional paper map. Early work on colour-coded maps has failed to find any support for their use (Christner and Ray, 1961; Williges and North, 1973). Spiker et al began by investigating which colours are easily confused in terms of luminances and chromaticity. Using this information, they developed a task in which subjects had to identify the presence or absence of target symbols appearing in one of three colours against 5 different background colours corresponding to position on the map e.g., water, terrain. Two control conditions used a homogeneous grey background and a luminance-matched grey background. Response time in the colour condition was always slower and the difference decreased with practice (0.2 secs). One reason for this might be that colour was overused and therefore no longer acted as a discriminating factor. We need to investigate the maximum number of colours which are useful but this may differ according to the task at hand. Colour might be most effective as a background to provide feedback about position. Alternatively colour coded targets against a neutral background might be useful. Too many colours lead to confusion e.g, colour coded targets against colour coded backgrounds.

5.1.3 Long Term Effects

Very few studies have looked at performance over longer periods of time. Christ and Corso (1983) studied a set of subjects over 9 months. They conclude that there are "no clear and consistent advantages for any one visual code set over others".

5.2 Empirical Studies

The studies reported above use various techniques which include search and identification tasks, employing mostly redundant colour and arbitrary symbols. From these studies the evidence suggests that in general, colour can improve search times over achromatic coding but identification times may not improve. Since identification plays the greatest role in many 'real world' tasks one should ask whether colour should be employed in such applications which do not require searching for targets. The task used in the study reported below differs from many of those presented in the introduction in that not only are we interested in non-redundant colour coding but how this information can be utilised in a qualitative manner to enhance the transfer of information. In addition, the stimuli (facial expressions) are unlike those used in previous studies in that they are more meaningful and less arbitrary symbols.

The study also differs methodologically. Subjects are required to 'familiarise' themselves with facial expressions first and then to associate the appropriate colours with the emotion expressed in the faces. Results from other identification studies suggest that non-redundant colour has no advantage over monochrome. In this study, it is proposed that colour will not add to processing time because a direct match can be made between colour and its associated meaning without decoding the facial expression. In a learning set, the absence of colour can only be effective when all the items can be successfully named. If items cannot be readily identified, the subject can make a good "guess" using the colour coding on the faces.

In the past, faces have been used as stimuli in a number of tasks. For example, they have been used by Chernoff (1973) to represent multivariate data. Sheehy and Forrest (1986b) and Sheehy et al (1988) have explored the notion of representing nonverbal communication at the interface in two ways. First, by permitting both user and intelligent front end to provide feedback to each other in terms of postures and facial expressions about how the interaction is proceeding, and, second, facial gestures of the user have been captured by an image processor so that the intelligent front end is able to respond to both the user's verbal and nonverbal gestures. Thus, the use of meaningful stimuli such as facial expressions have relevance to human computer interaction. A number of researchers have implemented this idea of representing qualitative changes through gestures or nonverbal communication. For example, changes in facial gestures have been used to represent severity of illness as well as changes in temperature in a medical setting. Another example similar to the work by Sheehy et al, is by SUN (DV-tools) who have used facial gestures to provide feedback on a user's progress. In a learning task in the domain of fire fighting, trainees learn how people respond in attempting to exit a building on fire. Different facial gestures are flashed on the screen according to the proximity and correctness of the moves made (TVDC/Chrysalis).

5.3 Pilot Study - Faces and Colour

5.3.1 Aims

The aim of this study is:

i) to test the validity of a number of faces by examining concordance among users' attributions of affect for different faces. These results are to be used in the main study reported later.

ii) to investigate the relationship between colour and emotion by examining the strength of commonly perceived associations between colour and affect.

5.3.2 Methodology

5.3.2.1 Subjects

Seventeen people (nine females and eight males) agreed to participate in this study. They comprised students, clerical and technical staff. All were native English speakers to avoid cultural differences in facial perception.

5.3.2.2 Materials

Eight faces depicting eight emotions were constructed by the author using the Microsoft Paintbrush software package. Twelve adjectives were generated as possible labels for the emotions depicted in the facial expressions.

5.3.2.3 Procedure

Subjects were given two sheets of paper: the first contained the eight faces and 12 adjectives. Subjects were instructed to examine the faces and write down the adjective which they felt best represented each facial expression. The same label could be used for more than one face, although this was discouraged. After completing the first part of the task subjects were asked to inspect their answers and make any changes. The second page contained the 12 adjectives and subjects were requested to write down a colour which they thought was strongly associated with or expressed that particular emotion (see Appendix 5.1).

5.4 Results

The eight most frequent responses for the eight facial expressions were: jealous, happy, angry, bored/afraid, embarrassed, grumpy, fed up and sad. For five of the faces (jealous, happy, angry, embarrassed, sad) the agreement between subjects about the appropriate adjectival label was

								
FIENDISH		3				3	1	
EMBARRASSED	16		1		1			
ANGRY						4		11
GRUMPY		2				5	1	2
BORED			6			1	4	
JEALOUS		11						
APPREHENSIVE		1	4				4	
FURIOUS						3		3
AFRAID			6				1	
SAD				16				
HAPPY	1				16			
FED UP				1		1	6	

Table 5.1 Facial Expressions and Associated Adjectives

	Fiendish	Embarrassed	Angry	Grumpy	Bored	Jealous	Apprehensive	Furious	Afraid	Sad	Happy	Fed up
Purple	6	2	2	2		3	1	6				1
Red		10	15	1				12	2		1	
Pink		11									2	1
Black	7		2	1	1				2	1		
Grey	1		1	9	6		2		1	3		5
Green	2				1	18	4		1			
Orange	1		2				2	1			3	
White					2		4		13	2		1
Blue	3	1		3	8		6			14		6
Yellow	1						1	1	2		16	
Cream					1							1
Brown				6	3		1			1		5
Mustard												1
Transparent										1		
Silver	1											

Table 5.2 Adjectives and Associated Colours

high (65%-94%). In cases where more than one adjective was associated, with equal frequency, with a particular colour the experimenter made the final decision. For example, 'grumpy', 'angry' and 'furious' were equally associated with the same facial expression. When 'angry' and 'furious' judgements were totalled they were greater than the number for 'grumpy' and 'furious' was selected as the label, as this adjective appears to be more 'extreme' and better defined (Table 5.1).

The most suitable colours were chosen in a similar manner: where two colours competed for the same facial expression, the next best was used. The colours chosen for the eight faces were: jealous - green, happy - yellow, angry - red, bored - grey, embarrassed - pink, furious - purple, fed up - brown, and sad - blue (Table 5.2).
The results of this study were used to investigate whether colour, when used in a qualitative manner could enhance subjects' responses on an identification task.

5.5 Study 3 - Faces and Colour

5.5.1 Methodology

5.5.1.1 Subjects

Ten male and eight female students participated for approximately 20 minutes each. All subjects were native English speaking volunteers.

5.5.1.2 Materials

The eight faces were reproduced on strips of card which

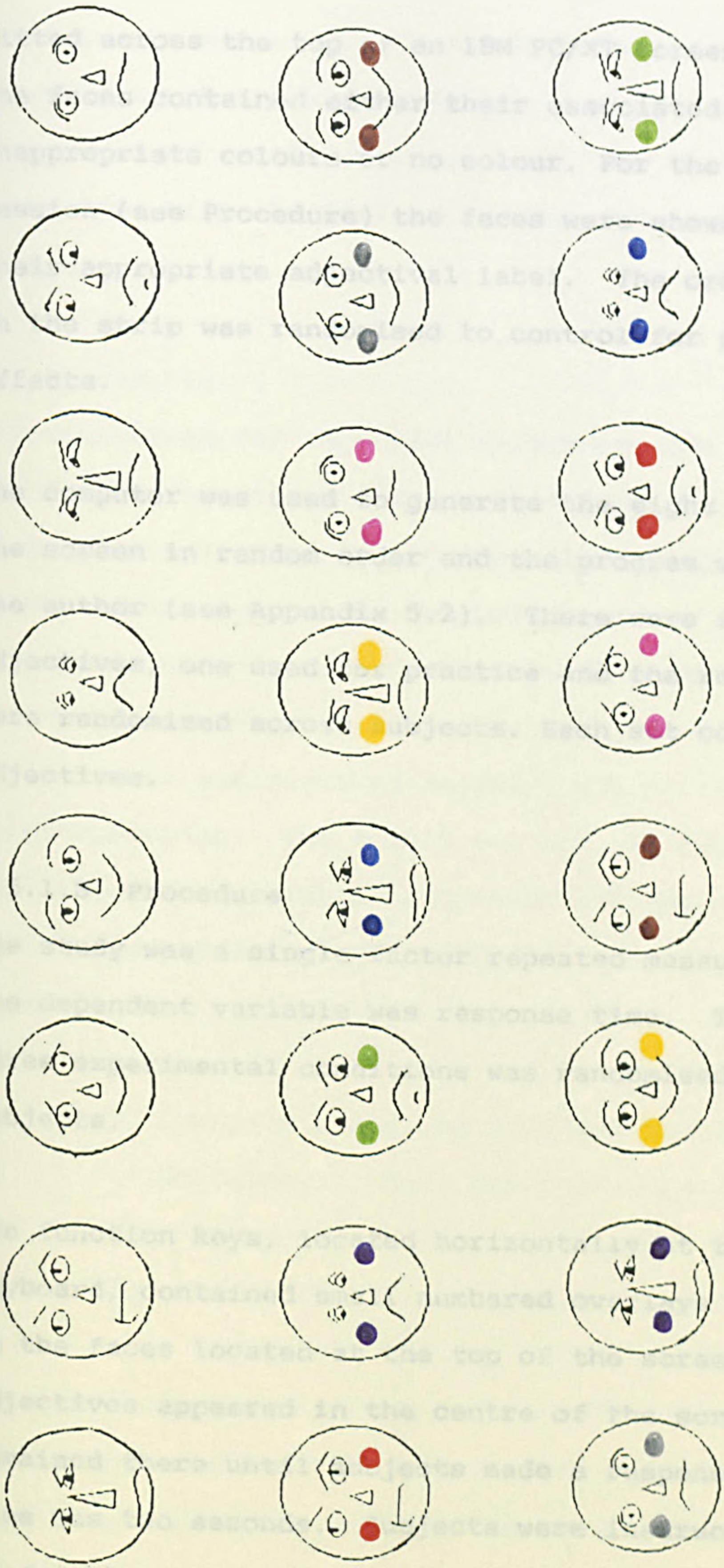


Figure 5.1 Materials: No Colour, Inappropriate Colour and Appropriate Colour-coded Faces

fitted across the top of an IBM PC/XT screen (Figure 5.1). The faces contained either their associated colours, inappropriate colours or no colour. For the practice session (see Procedure) the faces were shown together with their appropriate adjectival label. The order of the faces on the strip was randomised to control for position effects.

The computer was used to generate the eight adjectives on the screen in random order and the program was written by the author (see Appendix 5.2). There were four sets of adjectives, one used for practice and the remaining three were randomised across subjects. Each set contained 40 adjectives.

5.5.1.3 Procedure

The study was a single factor repeated measures design and the dependent variable was response time. The order of the three experimental conditions was randomised across subjects.

The function keys, located horizontally at the top of the keyboard, contained small numbered overlays corresponding to the faces located at the top of the screen. The adjectives appeared in the centre of the screen and remained there until subjects made a response. Inter-trial time was two seconds. Subjects were instructed to respond as quickly and accurately as possible to the appearance of the adjective by pressing the assigned key corresponding to the facial expression. Reaction times and errors were recorded by the computer. Subjects were given one practice session of 40 trials which allowed them to familiarize themselves with the different faces and their appropriate labels.

5.5.2 Results

Reaction times for incorrect responses were eliminated. Log10 transformations were performed on the raw data to eliminate curvilinearity in the data arising from a small number of extremely long or short reaction times. Table 5.3 shows the group mean values and standard deviations for the three conditions. A one-way analysis of variance with three levels and repeated measures was performed on the log transformations. The result was not statistically significant $F(2,34)=1.67$ and is consistent with previous work employing non-qualitative colour in identification tasks.

Table 5.3 Subjects Means and S.D. for the 3 Conditions

	No colour	Col. Approp.	Col. Inapprp.
$\bar{x}$	2.42	2.42	2.46
SD	.10	.12	.11

5.5.3 Discussion

It was expected that the qualitative use of colour would enhance performance, manifesting faster reaction times than when using non-qualitative or 'inappropriate' colour or no colour. Some subjects reported that the qualitative colour had helped while the non-qualitative, inappropriate colour produced a stroop-type conflict. Other subjects reported that the colour was not helpful and that they tried to attend to the position of the faces in the sequence.

Although statistically significant results were not found using traditional statistical techniques, the raw data demonstrate that more subjects were faster when colour was absent than when colour was present in either of its forms. In addition, more subjects were faster when qualitative, (i.e, appropriate) colour was presented relative to non-qualitative (i.e, inappropriate) colour (no colour=10 subjects faster; qualitative colour=6 subjects faster; non-qualitative colour=2 subjects faster).

Colour is, in general, processed in parallel and therefore should reduce the response time to identify colour coded targets over monochrome targets. However, in this study the colour may only have been adding to already meaningful stimuli so the greater identification times were in fact due to extra processing of colour. However, with practice, subjects in the colour condition should have performed better than subjects in the no colour and inappropriate colour conditions because it should be more difficult to process discrepant stimuli or stimuli with reduced information content (no colour) On the other hand, it is probable that after practice the subjects in the no colour condition could also change processing strategy and attend to the spatial arrangement of the faces on the screen.

The problem with using qualitative colour is that, for it to be effective as a technique to enhance performance, one must be able to associate readily the colour and object. A colour that is meaningful to one person may not be to another or, more simply, the colours may fail to induce the correct qualitative judgements because they are not sufficiently powerful. This is demonstrated in the study reported here by the wide range of judgements about the appropriateness of different colours for indicating different emotions. It appears that there are extremely few colours which have 'inherently appropriate' meanings, and outside these few there are likely to be strong learning effects.

The manner in which colour is used to improve performance and for what purpose are important. In some circumstances there may be no change in performance in terms of response time. Qualitative colour may simply act to give the user some information. For example, the use of colour to indicate temperature may not augment performance but simply inform the user about the "state" of a component. Study 2 demonstrates this to a certain extent with the use of non-qualitative colour. The colour (and shapes information) acted to inform the subject as to the position of a particular room.

In relation to adaptive interfaces, colour can be viewed as a support to the user, for example, the correct use of colour on display screens can aid search times but not identification times. In particular, the way in which colour is used both redundantly (e.g., to make the screen more comprehensible) and non-redundantly (e.g., red for stop) can be helpful to inexperienced computer users. Whether colour provides a greater value to the experienced user apart from its aesthetic value and personal preference has not been established or investigated. Finally, colour needs to be used carefully: a shade too much can destroy the potential effects.

5.5.4 Conclusions

The results of Study 3 suggest that performance times were not improved by the use of colour-coding intended to induce qualitative judgements. Although the identification task used here differs from those used in previous research, the findings are consistent. However, this may well be due to the position of the faces being the most powerful strategy working in the absence of colour. Thus, subjects no longer had to remember the actual facial expression but its position in a sequence. It has been suggested that some attributes such as position may well be automatically processed (e.g., Hasher and Zacks, 1979). Further work is required on qualitative colour while controlling for position effects.

Chapter 6 Case Study Designs: Methodology for Study 4

6.1 Introduction

Case study designs have rarely been used in the study of human computer interaction (cf. Mack, Lewis and Carroll, 1983). Differences in individual performance have been neglected and group performance on particular tasks has been the typical way of studying the user in human computer interaction (and used in Chapters 3 to 5 of this thesis). However, case studies are particularly important when we expect or predict that individual differences exist. Research with the object of creating adaptive interfaces from the viewpoint of the user's personal style (see Chapter 2) unfortunately has not led to in-depth study of the user on an individual basis. If we accept that there are individual differences in terms of cognitive and learning styles and personality characteristics which affect the way we handle problem solving tasks, then a next move is to study individuals who vary in psychological and learning dimensions and explore how these different characteristics affect performance. As important, is the need to observe how these characteristics interact to affect performance. Since it is impossible to study these interacting characteristics in a large group, these studies should essentially be single case or small n studies.

A second reason for using small subject numbers in human computer interaction research is that the nature of HCI entails that the interaction is one-to-one. That is, the aim of adaption is to personalise the interface to each individual user. Thus, the important factor is how the individual person interacts with the computer not how a group of individuals interact.

The next study in this thesis (Study 4) generated a wealth of both quantitative and qualitative data. The main problem associated with qualitative data and small n subject size is the analyses of the data produced. Although much of the data for all three subjects of the Study 4 have been grouped, the main thrust of the study lies in the notion that each individual's performance is reflected by his/her personal style. The problem of analysing small n studies is reviewed below.

6.2 Single-Case Subject Design

Intrasubject research is still not widely used or acknowledged as an adequate research design. Shapiro (1966) suggested that this is due to two assumptions. First, repeated measurement of the dependent variable on the single subject "produce(s) complex interactions which are difficult to interpret". Second, data are seen to lack external validity; that is, the results produced may not generalise to other cases. For this reason, the single-subject design requires replication using other subjects. However, as Shapiro points out, this assumption does not acknowledge that the important factor is identifying changes in the <u>relationship of events</u> and, furthermore, that it is possible to have repeatability of patterns of changes in two variables.

Shapiro (1966) wrote "the first step in the investigation of processes must logically consist of investigation in a number of individual cases" (p.5). In support of this, he cites Sidman (1952) and Bakan (1954) who proposed that data should not be combined from a number of individuals unless the differences are known to be due to experimental error. For example, the negatively accelerating curves of intellectual growth and learning have not yet been demonstrated in individual cases. Hayes (1981), reported "It is in the on-line clinical environment that the unique value of time series experimentation truly becomes apparent, yet little has been done to advance its use there" (p.194).

The two main areas in which single-case subject design and multiple baseline designs can have major impact are clinical psychology and educational psychology. The utilisation of repeated measurements of the dependent variable permit a wealth of data to be collected over many treatment phases, resulting in in-depth analyses of behavioural change (e.g. ABAB, A-B-A-B-BC).

The controversial issue surrounding these types of design is how the data should be analysed, by visual analysis or statistical procedures, or a combination of both. In the past, data have been analysed solely by visual inspection of graphs which depict changes in trends. However, these changes in trends are often subtle and are not detectable by "eyeballing" the data. Thus, visual analyses may not be appropriate when effects are small or when large variations in scores are present which may or may not be due to random chance (Hartmann, Gottman, Jones, Gardner, Kazdin and Vaught, 1980). Other disadvantages of employing visual analyses are, first, interpretation of the data is subjective and, second, there may be differences between analysts' interpretations. In fact, inter-evaluator reliability has been found to be low and independent of the evaluator's level of expertise (DeProspero and Cohen, 1979). On the other hand, the data are extremely easy to plot and explain (Parsonson and Baer, 1978).

The types of statistical techniques advocated for single-case subject (or small n) design are time series analysis and various types of analysis of variance.

One property of behavioural data is "serial dependency", that is, adjacent test items (or scores) tend to be more highly correlated with each other than scores not adjacent to each other. Jones, Vaught and Weinrott (1977) have found that serial dependency is common in behavioural studies - of 24 operant studies, 20 (83%) had significant lag 1 autocorrelations (autocorrelation refers to the amount of serial dependency existing between scores in the data). This is to be expected as people do not behave in a random manner (Jones et al, 1977). Serial dependency is particularly important when only a small n is used because a certain amount of variance will be due to this factor. The result is that total variance may be spuriously reduced and the null hypothesis falsely rejected (Type 2 error). The probability of Type 2 errors occurring in n=1 methodology is high while the probability of Type 1 errors is low (see also Baer, 1977).

The power of time series analysis is that it can account for serial dependency by transforming the raw scores into serially independent scores. Autocorrelations are performed which measure the amount of serial dependency between scores. So a lag 1 autocorrelation correlates adjacent pairs of scores while a lag 2 autocorrelation correlates scores 1 and 3, and scores 2 and 4 and so on. Conventional statistics such as analysis of variance and t tests are unsuitable because they work on the assumption that scores are independent and normally distributed, and that subjects are representative of the population. The use of such techniques will be discussed later.

Taking a number of published Journal of Applied Behavior Analysis (JABA) papers in which the authors had used visual inspection of the data to arrive at the conclusions of their studies, Jones, Wienrott, and Vaught, (1978) reanalysed the data using time series analysis. The studies varied in the number of scores for each phase (3-34 scores) and whether the number of scores was even or uneven. Statistically significant results were obtained for changes in levels and trends.

Jones et al (1978) further investigated the agreement between analysts using visual inspection and time series analysis, and whether this agreement varies as a function of the amount of serial dependency that exists between scores. JABA graphs were presented to 11 judges, all familiar with visual analysis (experience varied between 3 years and 17 years). Judges were asked to decide whether any reliable changes had occurred in trends in each treatment phase. The same data were subjected to time series analysis. Agreement was found to vary between chance and just 15 points above chance. When serial dependency was taken into account, agreement between visual analysis and time series analysis was found to be inversely related to the magnitude of the serial dependency of the scores. So, when serial dependency is low, mean agreement is relatively good between visual analysis and time series analysis (0.73), but drops to 0.54 when serial dependency is medium and still further to 0.5 when serial dependency is high. In addition, agreement was greatest (0.67) when the data were nonsignificant and poorest (0.52) when the data were significant. These results demonstrate that time series analysis should be used to supplement visual analysis (Jones et al, 1977, 1978) to prevent Type 1 errors occurring when visual analysis is used alone.

Alternative statistical techniques to time seriec analysis which have been suggested are various forms of the analysis of variance model. Behavioural data cannot be analysed by conventional statistics because they violate the assumption upon which the statistics were based (see Hays, 1963). The most important assumption is that scores must be serially independent of each other. A within cell dependency results in a positively biased F statistic (see Hartmann, 1974). Similarly, with t tests, autocorrelated data result in low standard deviations which inflate the t score and give highly inappropriately 'significant' results (McCain and McCleary, 1979). In addition, ANOVAs and t tests are inadequate because they operate on the basis of means and do not take into account any slopes or trends in the data (Peck, 1985).

Gentile, Rodin and Klein, (1972) advocated the one way fixed effects ANOVA model where treatment effects are used as between subject scores, and the number of observations as within subject effects. Their suggestion that this eliminates serial dependency has been heavily criticised. Kratochwill, Alden, Demuth, Dawson, Panicucci, Arntson, McMurray, Hampstead and Levin (1974) pointed out that most behavioural data are autocorrelated and that is supported by the Jones et al (1977) study mentioned earlier. Thus, the example of a 'typical design' of a behavioural study given by Gentile et al is thought to be atypical (Kratochwill et al, 1974). In addition, Hartmann (1974) has demonstrated that even the most perfect data set cannot be handled by the fixed effects ANOVA described by Gentile et al.

The study to be described in Chapters 7-9 produced a wealth of both qualitative and quantitative data. The employment of 3 subjects made interpretation of the data difficult, first, because traditional statistical techniques such as ANOVA could not be employed because of the criticisms given above. A multiple correlation was performed for some of the data but significance levels were difficult to obtain and there may have also been random effects of significance. The data had to be scrutinised for trends and patterns which were meaningful but this does eliminate the possibility of occurrence of both Type 1 and Type 2 errors.

It was not possible to use time series analyses on the data that involved ratings of emotions (e.g., tenseness, tiredness) because the number of data points was small.

One type of post-hoc analysis employed was to group subjects into similar and dissimilar behaviours (Chapter 9) in order to produce a category of "misunderstandings" experienced while learning to use the word processor. This type of analysis involves "objective" grouping of behaviours and is not expected to produce significant or non-significant findings, rather the result is a "dimension" which is then open to subjective interpretation. There are no right or wrong answers in qualitative procedures. Quantitative procedures also have weaknesses in that the design of the experiment is dependent on the types of analyses that can be later performed. Thus the subjective interpretation of the experimenter of qualitative procedures is no less contrived than the rigid design of experiments which will permit traditional statistical techniques.

Chapter 7 Personal Styles as Predictors of Performance

7.1 Introduction

Chapter 8 reports three case studies of novice computer users learning to use three different types of word processing packages. The aim of this chapter is to categorise users in terms of their personal style and then to predict performance in terms of the type of environment that would best suit him/her and the types of problems s/he is most likely to experience.

7.2 Types of Communication

Two decades ago the only form of human computer communication was via a terminal or printer with the computer and user taking turns in responding. This type of communication is often called half-duplex mode, that is, communication takes place using a single input-output channel. The introduction of new technology generated the concept of "multi-modal communication" which attempts to facilitate communication by increasing the input-output bandwidth.

Multi-modal communication permits the transfer of information on several levels - by speech, keyboard and pointing device. In addition, the use of graphics as a means of communication gives greater scope for understanding, in particular, for novice users. This can be achieved by way of using metaphors. Graphics can be viewed as a natural form of expression in human computer interaction in the same way as gestures and deixis is in human human communication (Bijl and Szalapaj, 1985).

Types of communication can be categorised functionally into the following:

i. Command mode or direct mode - the user is typically an experienced programmer or, alternatively, a frequent computer user. The user inputs commands (hence, "user-initiated") in a high level language which must conform to a particular syntax. The advantage of this type of dialogue is that data entry is fast (see Kidd, 1982), and the disadvantage is that the user needs to be accurate to avoid spending time correcting "pseudo errors". On the other hand, the emphasis in new technology is a move away from the traditional querty keyboard towards devices which make selection and input faster and easier.

ii. Menus - are a "computer-initiated" form of communication which involves a search through a database of menus. The system presents a list of choices on the screen to which the user responds by either typing in a number or positioning a cursor on the choice option using a pointing device. The advantage of menus is that memory requirements are reduced and they are ideal for novices and users who do not possess typing skills. The disadvantages include slower interaction especially if the list of choices is lengthy. Menu-based systems are not adaptive to the user so the more experienced the user becomes the more frustrating s/he will find the interaction. One solution is to allow the user to respond as soon as the appropriate choice has appeared on the screen. For the novice user, multi-level menus may be confusing, hence good on-line documentation is advisable.

iii. Form-filling - is another form of "computer initiated" communication. The screen becomes a form which the user fills in. An advantage of form-filling is that it is based on a user's knowledge about filling in forms so there is virtually no training needed. The system can be slow but this can be remedied by allowing the user to enter the relevant information separated by commas and by-passing the cursor movement involved.

Menus and form-filling are prompt type dialogues, that is, the computer prompts the user for a response. Frequently, these choices are given in order that the interaction itself can be well structured.

iv. Iconic-based systems, such as those that use the "desktop" metaphor, involve screen representation of familiar working environments. For example, objects typically found in the office (files, in and out trays, clocks, wastepaper basket) are represented graphically, often by means of "icons". These can be manipulated and dragged across the screen using a pointing device. For example, a mouse slides across a desk or flat surface, and registers as a movement on a grid which in turn moves the cursor on the screen proportionately. Files can be opened and closed by placing the cursor on the icon representing the file and clicking a button situated on top of the mouse.

Iconic-based systems are considered one of the easiest with which to interact, combining keyboard skills and graphical interaction. They are especially designed with novice computer users in mind, but also allow some amount of tailorability such as dispensing with the icons and using textual labels.

The types of dialogue modes mentioned differ in the amount of support the user is given during the interaction. Some types of dialogue such as command mode are extremely flexible while others tend to be more structured.

7.3 Word Processing Packages

The three word processing packages employed in the study described in the next chapter were: Word Star, Word Perfect and Perfect Writer II. Each employs different communication modes and thus they necessarily differ in the amount of environmental support given to the user.

All 3 word processors have one factor in common: they all use menus to some extent. Despite this, they can be viewed as operating in quite diverse ways.
Word Star (WS) employs a type of command language. The user is offered a combination of the control key and letters (e.g. ^QD moves the cursor to the end of the line).

There are three levels of difficulty depending on how familiar the user is with WS. Subjects in this study experienced the lowest level of difficulty: menus remained on the screen showing the different commands for different operations. However, subjects needed to know which menus contained specific operation commands. There are also other commands known as "Dot" commands which send commands directly to the printer. These commands are not found on the menus.

WS is not generally thought to be an easy word processor to use because the sequence of commands are not mnemonically based. In addition, communication is mostly user-initiated, that is, the user leads the dialogue and the computer responds.

Word Perfect (WP) employs a prompt type language. Users perform operations by pressing function keys (with shift, control or alt keys). Sometimes this alone will produce the desired effect but mostly the computer presents the user with a menu of options and prompts the user to respond.

WP is to a large degree computer-initiated, that is the computer leads and the user responds. However, the package is not as restrictive as pure menu-driven dialogues. For example, the user can ask the computer to reveal any codes used, such as tabs and highlighting, and then edit these codes.

Perfect Writer (PW) in its 'normal state' is menu-driven. For the majority of the interaction, the user chooses options from a screen menu. This type of interaction is entirely computer-initiated, restricting the user's moves. There are function key commands for the more experienced users. Subjects were discouraged from using these commands although it was faster to use the function keys than following the menus provided (see Appendix 7.1 for examples of screens).

Subjects are expected to encounter more difficulties with WS than WP and PW. PW, in general, is expected to be the least difficult with which to interact. This general formula may be modified by a user's personal style.

7.4 Personal Styles

The types of cognitive and learning styles and personality inventory chosen were:

Reflectivity-impulsivity (MFFT)
Field dependence-independence (EFT)
Divergent thinking
Operation-comprehension learner (Styles of Learning Questionnaire)
Introversion-extraversion (EPI)
Stable-unstable (EPI)
(see Appendix 7.2).

The use of case study design is problematic. The lack of relevant and meaningful norms for the tests with which to view the three experimental subjects resulted in a collection of new norms from other students.

14 students completed the full battery of tests. For the MFFT, and Divergent thinking tests, norms were collected from 14 students (7 female, 7 male) and norms from 33 students (16 female, 17 male) for the EFT. Despite the small subject size, a median split was used to categorise the subjects. 14 students also completed the Styles of Learning Questionnaire. With the addition of the three experimental subjects, 17 people had completed the full battery (see Appendix 7.3 for subjects' norms). Their data were analysed using a Spearman's rho.

7.4.1 Learning Styles

The Learning Styles Questionnaire has already been validated by Entwistle and colleagues. The Learning Styles Questionnaire consists of four dimensions: these are meaning orientation, reproducing orientation, achieving orientation and holistic orientation. Entwistle's study on over 2000 students demonstrated how the scales and subscales correlate with each other. He found that some of subscales had better intercorrelations and the new dimensions became meaning orientation (deep approach, comprehension learning, relating ideas and use of logic and evidence) reproducing orientation (surface approach, improvidence, fear of failure and syllabus bound), achieving orientation (achievement motivation, intrinsic motivation and strategic approach) and non-academic orientation (deep approach, negative attitudes and globetrotting). Subject's scores for the above, in addition to scores for holistic orientation (score for holist style + pathology + reversed score of serialist style and pathology), holistic style (holist + globetrotting), serialist style (serialist + improvidence), versatility (holist + serialist) and pathologies (globetrotting + improvidence) were intercorrelated using a Spearman's rho correlation. In line with Entwistle's results, there were significant correlations between meaning orientation and holist style (0.89, $p<0.01$), holistic orientation and holist style (0.56, $p<0.05$) and holistic orientation and versatility (0.69, $p<0.01$), and between achievement motivation and versatility (0.65, $p<0.01$). Another dimension was also prominent: there were significant positive correlations between reproducing orientation and serialist style (1.00, $p<0.01$) and non-academic orientation and pathologies (1.00, $p<0.01$). There was also a trend towards a positive correlation between serialist style and pathologies but this did not reach statistical significance (0.43). These results suggest that it is possible to group subjects on learning style using the Learning Styles Questionnaire.

7.4.2 Battery of Tests

Mean times to respond in the MFFT correlated positively with holistic orientation (0.49, $p<0.05$) and holist style (0.59, $p<0.05$) suggesting that reflectives tend also to be comprehension learners. Mean times to respond were also related to the number of inaccurate responses demonstrating that reflectives tend to be more accurate than impulsives (0.63, $p<0.01$). A trend was found to exist between MFFT response times and divergent thinking but did not reach statistical significance (0.44). There may be a tendency for reflectives to be more divergent in their thinking strategies than impulsives.

Reproducing orientation and divergent thinking correlated negatively indicating that people on this dimension do not think divergently (0.46, $p<0.05$). Other significant correlations included instability and reproducing orientation (0.67, $p<0.01$), and non-academic orientation (0.67, $p<0.01$) and serialist style (0.68, $p<0.01$) and pathologies (0.67, $p<0.01$). Reproducing orientation, non-academic orientation, serialist style and pathologies are interrelated factors and have been found to cluster by Entwistle.

The expected correlation between the EFT and the MFFT, in particular the incorrect scores did not emerge.

Given the very small subject sample size, these results are encouraging, suggesting that the relationships are fairly robust and not the product of using large samples. The intercorrelations for the styles of learning questionnaire which have emerged are consistent with the findings reported by Enwistle et al on over 2000 students. These results permit classification of subjects and are a good basis from which to explore, in depth, the relationship between style and performance in the domain of human computer interaction reported in the next chapter.

From the knowledge gained through theory and research on cognitive and learning styles and personality characteristics, it has been possible to construct general descriptions of each attribute. The results from the multiple correlation matrix of the tests have also contributed to the descriptions. These descriptions are later be used to predict subjects' performance on each word processor.

Impulsivity: Impulsives are more likely to make errors than reflectives. They will probably need a large amount of support and become bored quickly especially if the task is difficult (Van der Veer et al, 1985). Given the choice between learning from a manual and learning from an "expert", impulsives are likely to prefer the latter.
Reflectivity: Reflectives, in general tend to favour accuracy over speed. They may also show signs of divergent thinking therefore they are likely to become bored if the task is easy and routine. Reflectives will probably need less support than impulsives and prefer to learn from a manual.

Field-independence: Field independent subjects can work in a relatively unsupportive environment. They tend to perform well on a variety of word processors (Fowler et al, 1986).

Field-dependence: Field dependent subjects are driven by extrinsic factors and therefore need a supportive environment in which to work. They may work on a trial and error schedule (Fowler et al, 1986).

Divergence: Divergent thinkers tend to perform well on complex and diverse systems. They become bored using routine procedures especially if they are long and tedious.

Serialist style (Operation learners): Operation learners work from basic details, transforming descriptions into actions. They do not benefit from large comprehensible manuals and therefore tend to be more suited to simply operated packages. They may also have a tendency to instability.

Comprehension learners: They work from general information about how the package works, making use of analogies. They continuously attempt to construct a model of how a software package operates. They are likely to perform well on both simple and complex packages. Comprehension learners tend also to be reflective.

Versatile learners: These learners are the most successful. They are able to adopt the correct strategy to obtain the most efficient result.

Introversion: They need a fairly supportive environment which does not make them over-anxious.

Extraversion: Extraverts prefer variety and become bored with routine procedures.

Unstable: Unstable subjects tend to work with trepidation. They may work better with a simply operated word processor.

Stable: Stable subjects rely on their own judgements and from the instructions in the manual.

The above gives a list of descriptions of how typical users of each category may behave. None of the categories should be viewed as mutually exclusive; rather each characteristic will modify the user's behaviour. Those characteristics which cluster together best will influence a user's general behaviour.

7.5 Aims

The aim of this section is to describe the usefulness of cognitive and learning styles and personality factors in predicting users' performance on three word processors varying in mode of communication and ease of use. Chapter 8 describes how the three experimental subjects perform in vivo on the word processors.

7.6 Predictions

An attempt is made below to predict from each subject's unique characteristics, the word processor most suited to him/her. This is achieved by matching each characteristic with the word processor that supports that characteristic and then observing which word processor is most dominant.

7.6.1 Subject 1, Age: 24 years, Sex: female

Characteristic	Word Processor
Impulsive	PW
Field Dependent	PW/WP
Convergent Thinker	PW/WP
Versatile Learner	WS
Extravert	WS
Unstable	WP

Prediction = Perfect Writer or possibly Word Perfect

Subject 1 requires a word processor offering a supportive environment. She prefers to learn from another person rather than consult the manual. Therefore, she prefers manuals that are simple, giving basic details of how to begin rather than in-depth comprehensive manuals.
She is more suited to a computer-initiated mode of communication because she shows little flexibility in her problem solving approach. She may become bored using a word processor that utilises routine operations but this is preferable to a word processor which is too complex, which she will find to be more boring. In addition, she may not complete her learning of the word processor if left unaided.

7.6.2 Subject 2, Age: 24 years, Sex: Male

Characteristics	WordProcessor
Impulsive	PW
Field Independent	WS
Convergent Thinker	PW/WP
Versatile learner	WS
Extravert	WS
Unstable	WP

Prediction= Word Star

Subject 2 attempts to understand why and how the system operates, constructing a model in his mind. He makes his moves logically but also show signs of trepidation. He is able to work within a fairly unsupportive environment but has a tendency to become bored if the task is too complex. In an environment supported by an "expert", he will indulge in a great deal of interaction with the "expert" in an attempt to clarify any misunderstandings. Subject 2 likes variety but is not able to think divergently.

7.6.3 Subject 3, Age: 23 years, Sex: Female

Characteristics	Word Processor
Reflective (inaccurate)	WS
Field Dependent	PW/WP
Divergent Thinker	WS
Comprehension Learner	WS
Extravert	WS
Stable	WS

Prediction= Word Star

Subject 3 tends to become bored if the task is too easy or repetitive. She will prefer to work with a more interesting and flexible type of word processor to which she can readily adapt. Subject 3 is able to use many sources of knowledge to think up alternative ways of dealing with problems but is also prone to filling in too many of the details without sufficient evidence. She prefers to work by consulting the manual and spending time constructing a model of how the word processor operates before interacting. She can work well unaided by the "expert" but may become hindered by her own inaccuracies.

7.7 Summary

An attempt has been made to define what types of behaviour are associated with each personal style, in order to predict the type of environment each of the three experimental subjects will best perform. Since no other research of this kind has previously been conducted, the predictions can only be validated through observation of the subjects' actual performance. Chapter 8 describes a study in which the categorised subjects learn to use three different types of word processing packages.

Chapter 8 Can Personal Style Predict Performance? Three Case Studies

8.1 Introduction

This chapter reports three case studies of novice computer users learning to use three different types of word processing package. Each word processor differs in the environment it provides. Chapter 7 has provided operational definitions of each personal style and how they are expected to affect performance strategies. It is intended that the case studies should demonstrate the usefulness of categorising types of user in terms of their particular cognitive and learning style and personality characteristics. These categorisations are then used to predict performance, that is, how the users handle and solve problems, and how they interact with the environment the word processor provides. In addition, performance is also investigated through the use of questionnaires which provided information such as attitudes and motivation as well as emotion (tenseness, sadness, tiredness and boredom). It was thought that the questionnaires would help establish the intricacies of the interaction.

8.2 Methodology

8.2.1 Subjects

One male and two female postgraduate students (of similar intelligence, as measured by AH4 parts I and II) ranging from limited previous use of word processors to no previous use of word processors. All subjects had some familiarity with statistical analyses on computers. Again this ranged from mere data entry to more sophisticated use of statistics. None of the subjects had used an IBM PC and none had any experience or familiarity with the word processing packages employed in this study. Subjects received payment for approximately 18 hours of participation over 3 weeks.

8.2.2 Categorisation of Users

Before commencing the learning part of the study, subjects completed a number of tests. These were:

Cognitive and Learning Styles - The Embedded Figures Test
The Matching Familiar Figures Test
A Learning Styles Questionnaire

Thinking Strategy - Divergent Tests

Personality - Eysenck Personality Inventory (Form B)

All tests were conducted on an individual basis.

8.2.3 Word Processing Packages

For details see section 7.3

8.2.4 Instructions

Subjects were instructed that they would be learning to use 3 word processing packages and that during the sessions they would be required to "think aloud". They were asked not to attempt to memorise commands. By "thinking aloud", they were to describe what they were doing and thinking. This included unexpected responses from the computer, what they thought was going on and explaining the responses of the computer.

8.2.5 Procedure

The study lasted 3 weeks: subjects were tested on 3 consecutive days each week on one of the 3 word processors. There were occasions when a one-day interval occurred through illness. This should have a minimal effect on performance. Each subject learned one of the word processors each new week. Although the study primarily consists of individual case studies, the order of introduction to the word processors was rotated. All the sessions were video-recorded for later analyses.

On day 1, the subject learned some basic operations e.g., editing text and highlighting text (centre, bold, underline and capitalise). More advanced skills were learned on day 2 e.g., replace and move text, formatting text and file manipulation. A more full account of the operations learned can be found in Appendix 8.1. On day 3, all the skills they learned in the first two sessions were tested without the aid of a written manual. Help from the 'expert', who was present throughout each session, was only given if the user could not recover from an error which would have prevented the study continuing. Each session lasted up to two hours and only occasionally did subjects exceed two hours.

8.2.6 Materials

Subjects were required to type in a piece of text approximately one page long. The text was an extract from learning material in the WP manual (see Appendix 8.1). At all times they had copies of the text before and after the corrections and modifications had been made.

8.2.7 Manuals

Manuals for the three different word processing systems differed in their organisation, detail and completeness. This was an unavoidable factor and it is possible these differences influenced the results to some extent. Subjects were given a summary of commands for WS, the reference section of WP and edited chapters of the PW manual. When manuals were withheld on day 3, subjects were not without some support: they could consult the menu of commands located at the top of the screen for WS. A template of commands surrounded the function keys for WP and for PW, a tree diagram depicting routes through the menus was supplied.

8.2.8 Questionnaires

The author designed a number of questionnaires which would provide a variety of information which might affect the subject's performance. It was hoped that gauging these types of information would help construct a full account of what was happening during an interaction.

Before each session, the subject filled in two questionnaires designed to elicit their attitudes towards computers and how motivated they were to learn how to use word processors. These questionnaires were Likert type scales using 7 points. Visual analogue scales for rating emotion (tiredness, tenseness, sadness and boredom) were also given. Half-way through the session the subject was required to fill in the emotion scales again. At the end of each session, the two questionnaires and emotion rating scale were completed (see Appendix 8.2 for questionnaires).

After the experimental sessions of days 2 and 3, subjects completed visual analogue scales to rate the degree of difficulty experienced on particular task components e.g. editing, formatting, highlighting. These scales give some indication of how easy the subject found the different word processing packages to use.

Visual Analogue Scales (VAS) have become a popular technique to record mood change in subjects. Typically, they constitute 10cm horizontal lines with two contrasting or opposing words anchored at both ends. VASs are thought to be more sensitive than fixed point scales (FPS) as FPSs may have insufficient points to demonstrate subtle changes (Hill, 1987). In addition, it has become traditional to use unipolar VAS e.g., not bored-very bored to avoid a skewed effect towards the positive end of bipolar scales e.g., tired-alert (Bond and Lander, 1974).

The impetus for using questionnaire techniques is to monitor the intrinsic factors of the user when interacting with the computer. There is some evidence that poor motivation and negative attitudes have an adverse effect on performance (e.g., Breakwell et al, 1986). Additionally, the subject's mood on a particular test day may also affect performance. The emotion rating scales cover these possibilities.

8.2.9 Verbal Protocols

During days 1 and 2, verbal protocol techniques were employed to provide an inventory of types of misunderstandings occurring during the interaction. Subjects were required to "think aloud", that is, to explain what they thought was going on and to indicate if the computer did not do what they predicted, and why. It was hoped that this could contribute to the construction of an outline model of what the user understands about the word processing package. At the end of each session, the subject explained what s/he had learned about the word processor. This included information about what s/he saw as the structure of the communication mode and how it operated. Subjects were also invited to give their own personal views and preferences for a particular word processor.

8.3 Results - Quantitative Analyses

The study generated approximately 40 hours of video recording. A preliminary analysis of all the tapes was made. This involved counting the number of keystrokes that subjects took to complete the different parts of the task. These were: insert character, insert space, insert word, insert line, delete character, delete spaces, delete word, delete sentence, concatenate lines, centre, bold, underline, capitalise, move, replace, flush right, justify, double space, page number, change margins, reformat, rename and print. In order to obtain a normalised score, the number of keystrokes taken by the subject to achieve an action was divided by the minimum number of keystrokes that could be used to achieve that action. There were some features which not all of the word processing packages possessed. It was decided to eliminate these from the data in order to obtain comparability across word processors. The following features were removed: capitalise, flush right, and reformat. In addition, keystrokes for boldfacing were also removed because the description in the manual was misleading. The mean number of keystrokes for each subject for the two sessions are shown below. Session 1 refers to days 1 and 2 and session 2 to day 3.

Irrespective of the user's personal style and personality characteristics, PW appeared to be the easiest with which to interact in terms of the number of keystrokes taken to complete each task. The mean for all subjects is 2.31 (a mean of 1 being a perfect score). Subjects 2 (S2) and 3

(S3) were able to improve in the second session, while Subject 1's (S1) performance decreased. However, the score for S1 for session 1 was already extremely good (1.53) (see Table 8.1).

Performance in terms of number of moves was comparatively similar for WS and WP, 2.99 and 3.19 respectively. A closer look revealed that subjects performed better on WP than on WS in session 1 but the reverse was true for session 2. The reason for this may be due to the differences in conditions for session 1 and 2. Namely that, with all things being equal for session 1, performance was superior for WP, but for session 2, when manuals were withdrawn, subjects were at an advantage using WS over WP. This appeared to be due to the better on-line help facilities for WS. PW is menu-driven, so again it was less likely that problems would arise when the manual was withheld.

It was predicted that subjects would find WS to be the most difficult of the word processor to use, followed by WP, and then PW. S1 and S2 performed according to this prediction: they used more moves for WS than for WP and PW, and more moves for WP than PW. For S3, performance on all 3 word processors were similar for session 1. For session 2, her performance improved for PW (3.89 to 2.11) and slightly for WS (3.57 to 3.21) but decreased for WP. These results are consistent with the notion that the environments offered by the word processors influenced the ability of the subjects to operate them. S3 was the least experienced of the subjects and the data suggest that she found all the word processors equally difficult.

Table 8.1: Mean number of moves (for 19 features)

	Word Star	Word Perfect	Perfect Writer
Subject 1			
Session 1	3.15	2.8	1.53
Session 2	2.81	2.7	2.07
Subject 2			
Session 1	3.11	2.95	2.39
Session 2	2.09	2.5	1.87
Subject 3			
Session 1	3.57	3.56	3.89
Session 2	3.21	4.65	2.11
Total	17.94	19.16	13.86
Mean	2.99	3.19	2.31

S1 and S2 performed fairly similarly for session 1 (with the exception of S1's superior performance on PW). However, for session 2, S2 shows the greatest amount of improvement on all three word processors. In particular, S2's performance on WS is much better than the other two subjects. Overall, S3's performance is poorer than S1 and S2.

The number of moves can only be used as a measure of performance if it reflects the output of the subject's sessions, that is, the accuracy and completion of the different features of the task. In addition, it is necessary to take into account the amount of help given to the subjects. The 9 most complex features (see Table 8.2)

Table 8.2 Performance Data Showing Accuracy and Completion of Tasks

	Word Star		Word Perfect		Perfect Writer		
	Day1	Day2	Day1	Day2	Day1	Day2	
Subject 1							
Flush R.	2	2	2	2	2	2	
Bold	2	2	2	2	0	0	XX
Underline	2	2	2	2	2	2	
Capitals	2	2	2	2	2	2	
Double Sp.	0	0	2	2	2	2	
Page No.	2	2	2	2	2	2	
Centre	1	2	2	1	2	1	
Justifn.	2	2	2	2	2	2	
Replace	2	2	2	2	2	2	
Move	2	2	2	2	2	2	
Total	15	16	18	17	18	17	
Subject 2							
Flush R.	2	1	2	2	2	2	
Bold	2	2	2	2	0	0	XX
Underline	2	2	2	2	2	2	
Capitals	2	2	2	2	2	2	
Double Sp.	2	2	2	2	2	2	
Page No.	2	0	2	0	2	2	
Centre	2	2	2	2	2	2	
Justifn.	2	1	2	2	2	2	
Replace	2	2	2	2	2	2	
Move	2	2	2	2	2	2	
Total	18	14	18	16	18	18	
Subject 3							
Flush R.	2	2	2	2	2	0	
Bold	2	2	2	2	1	0	XX
Underline	1	2	2	1	2	2	
Capitals	2	2	2	2	1	2	
Double Sp.	2	2	2	2	2	2	
Page No.	0	2	2	2	2	2	
Centre	2	0	1	2	0	0	
Justifn.	2	2	2	2	2	2	
Replace	2	2	2	2	2	1	
Move	2	2	2	2	2	2	
Total	15	16	17	17	15	13	
Total	94		103		99		

were scored according to the accuracy and completeness of the work each subject produced at the end of each session.

For each feature two points were given for total accuracy and completeness, 1 if the feature had been completed but not accurately, and 0 if the feature had not been accomplished.

Although subjects showed very little improvement for WP, performance in terms of accuracy and completeness was better for WP than any of the other two word processors. Accuracy for PW was the lowest, but this was mainly due to S3's poor performance. This may possibly be due to the fact that PW was the first word processor for S3 in this study (order was rotated for the subjects). S1 and S2 performed as well on PW as they did on WP and superior to WS.

8.3.1 Ratings of Difficulty

In terms of the perceived difficulty of 6 general components of the task (cursor movement, editing, highlighting, formatting, printing and file handling), WS was perceived by subjects to be in general more difficult ($x=42.46$) than WP ($x=35.39$) and PW ($x=34.86$). Difficulty ratings did not appear to reflect increases in tenseness; that is, increased difficulty did not result in the subject feeling more tense.

8.3.2 Motivation and Attitudes Questionnaire

Motivation and attitudes only fluctuated by 1-2 points on the 7 point Likert scale during the sessions and therefore do not appear to reflect performance directly.

8.3.3 Ratings of Affect

There does not appear to be any clear cut relationship between ratings of boredom and tenseness. In the case of S1, there is a trend towards an inverse relationship; that is, as tenseness increases, boredom decreases. In the case of S2, boredom increases with increases in tenseness. There appears to be no linear relationship between tense and bored ratings for S3. In addition, the ratings do not appear to reflect a linear relationship with performance on any of the word processors.

8.3.4 Anxiety and Performance

According to Eysenck (1970) introverts have a lower threshold for anxiety than extraverts. A high level of anxiety is thought to affect performance by focussing attention on irrelevant detail. The EPI was completed by all the subjects but none obtained a score to permit a classification as "introvert". On the whole, anxiety did not appear to affect performance.

8.3.5 Extraversion and Ratings of Tenseness

Since all the subjects were "extraverts", no firm conclusions can be drawn about differences in anxiety between introverts and extraverts. However, there appears to be a trend depicting a relationship between extraversion and the degree of self-reported tenseness. S2 was the least extravert of the 3 subjects and rated his feeling of tenseness as being much higher than the other subjects. Similarly, the increases in tenseness experienced during the interaction were also very high. On the other hand, S3, who was the most extravert, rated her feeling of tenseness as much lower than the other 2 subjects. The accumulative scores for tenseness are shown below in Table 8.3. Accumulative increases refer to changes in increases of particular emotions during the interaction rather than the total of all the mood ratings over the session.

Table 8.3: Accumulative Increases of Tense Ratings

		Word Star	Word Perfect	Perfect Writer	
S1	Day 1	0	7.5	13.5	
	Day 2	23.0	0	3.0	
	Day 3	0	25.0	6.5	
	Total	23.0	32.5	23.0	= 78.5
S2	Day 1	33.5	5.0	8.5	
	Day 2	42.5	14.5	45.0	
	Day 3	17.0	0	4.0	
	Total	93.0	19.5	57.5	= 170.0
S3	Day 1	0	1.5	4.0	
	Day 2	0	5.0	5.0	
	Day 3	1.0	25.0	22.5	
	Total	1.0	31.5	31.5	= 64.0

8.3.6 Ratings of Difficulty and Boredom

Intuition suggests that there exists an important interaction between difficulty and boredom. A task which is too easy and a task that is too difficult may have very similar effects: they may both lead to boredom. In the case of a task being too difficult, the subject may even "give up". In general, these suggestions were borne out in this study. WS had previously been judged to be the most difficult of the word processors and PW the easiest with which to interact. The largest increases of boredom are found when subjects interacted with WS, and fairly large increases are found with PW (see Table 8.4 below).

Table 8.4: Accumulative Increases in Bored Ratings

		Word Star	Word Perfect	Perfect Writer		
S1	Day 1	52.5	17.5	12.0		
	Day 2	45.5	2.5	27.0		
	Day 3	18.5	8.0	12.0		
	Total	116.5	27.5	51.0	=	195
S2	Day 1	37.0	3.5	11.0		
	Day 2	42.5	0	24.5		
	Day 3	23.5	0	9.0		
	Total	103.0	3.5	44.5	=	151
S3	Day 1	1.0	0	3.5		
	Day 2	16.5	6.5	50.5		
	Day 3	11.0	5.0	7.0		
	Total	28.5	11.5	61.0	=	101
	Total	248.0	42.5	156.5		

8.3.7 Ratings of Boredom, Impulsivity and Difficulty

Increases in boredom may be related to impulsivity. The data suggest a positive relationship between ratings of boredom and impulsivity; that is, the more impulsive a user, the more likely the user is to become bored when the task becomes difficult. S1 and S2, who can be categorised as "impulsive" using the median split, showed the greatest increases in boredom when interacting with WS. S3, who was found to be "reflective" (but inaccurate) on the median split measure, showed greater increases in boredom when the task was too easy as when interacting with PW. All subjects showed the smallest amount of increase in boredom while interacting with WP.

8.3.8 Ratings of Sadness and Tiredness

The purpose of these scales was to account for any unusual deficit in performance. For example, if a subject was upset or particularly tired on a study day, performance might be reduced.

If a rating of sadness or tiredness is high at the beginning of the session, the probability that further increases occurred during the interaction was less than if the rating was initially low. Thus, it is essential to take into account the first rating and then any changes occurring after that (see Appendix 8.3 for the accumulative increases in sadness and tiredness ratings).

High ratings for sadness and tiredness and large increases did not appear to interfere with their overall performance. For example, if the ratings and performance scores are compared, one finds that subjects might actually perform better on high sadness and tiredness days. Lack of an interaction between affect and difficulty eliminates the possibility that the degree of affect is modified by the difficulty of the word processor. In sum, the ratings do not appear to reflect performance either in terms of completion of the tasks or in terms of number of moves taken.

8.3.9 Summary

In general, subjects performed best in terms of number of moves using PW and worst using WS. Performance, in terms of number of moves and completion of tasks, did not appear to be affected by motivation and attitudes as measured by the 7 point Likert-type scales or moods as measured by the VASs. Some patterns did emerge, such as interactions between ratings of difficulty, boredom and and impulsivity. It appears that a task that is too difficult or too easy leads to boredom which is modified by the reflectivity-impulsivity dimension. That is, impulsives become bored when the task is too difficult while reflectives become bored when the task is too easy. These results are consistent with the defined category of impulsivity in Section 7.4.2.

8.4 Behavioural Analyses

A fuller analysis of the tapes was made for session 1 (days 1 and 2) by coding a number of behaviours in real time. The tapes were coded by the author and no attempt was made to account for inter-rater reliability since the measures are discrete and therefore can be coded accurately. Frequency counts for the following behaviours were made: consulting the manual; consulting the help menu; expert explains; expert advises; novice asks a question; and novice asks for help. Time on behaviour was recorded for consulting the manual, consulting the help menu and overall time to complete a task.

Operational Definitions:

Expert explains: explaining to the subjects why something occurred.

Expert advises: prompting the subjects or giving advice as to how to proceed.

Novice asks for help: refers to both direct and indirect asking; it also includes "how do I do this?" questions.

Novice asks a question: includes all other questions relevant to interacting with the computer; for example, "can I do this?"

The frequency counts for consulting the manual and consulting the help menu were collapsed, as were their respective amounts of time spent in consultation. This was because the help menu for WS was an integral part of the package and was used by all the subjects. Only one subject attempted to use the help menus for the other two word processors.

The time to format the document after changing the margins was excluded from the overall time to complete the tasks because two of the word processors formatted automatically on-screen. Two of the subjects experienced great difficulty with this feature on the remaining word processor resulting in an extremely long time on task, which might in turn bias the results.

Four of the behaviours (expert advises, expert explains, novice asks a question, novice asks for help) were coded using an event recorder, while the remaining two or one depending on the word processor, were recorded by hand concurrently. Frequency tables and time spent on a particular behaviour are shown in Tables 8.5-8.7.

The use of a small subject design negated the possibility of employing many of the traditional statistical tests (see Chapter 6). It was, however, possible to enter all of the data into a Pearson's Moment correlation and obtain a correlational matrix. These data included not just the frequencies and times produced by the behavioural analyses but also the data produced by counting the number of keystrokes used for each word processor. In addition, subjects' personal style scores were included. This resulted in a 3x32 correlational matrix.

The employment of a small number of subjects resulted in difficulty reaching significance levels (therefore Type 2 errors are very high) but at the same time, Type 1 errors might also occur from spurious correlations. To attain a 5% level of significance, it is necessary to have a correlation of 0.997 and for a 1% level of significance, a correlation of 0.999. For this reason, some correlations which did not reach statistical significance are highlighted and discussed. Correlations which are non

	Consults manual/menu	E advises	E explains	N asks Q	N asks help	Time on manual/menu	Time on task
C1							
T1	-	13	6	15	9	-	22.53m
T2	4	10	8	18	15	1.32m	41.36m
T3	6	3	3	6	1	4.00m	18.36m
T4	18	5	1	19	6	10.46m	51.46m
T5	7	4	-	18	2	5.38m	22.48m
Total	35	35	18	76	33	21.56m	*135.44m
C2							
T1	-	-	-	2	-	-	9.19m
T2	8	2	-	4	2	6.45m	15.16m
T3	9	2	-	2	-	6.19m	14.28m
T4	31	2	-	10	4	6.20m	23.55m
T5	18	8	1	12	6	10.21m	20.41m
Total	66	14	1	30	12	29.45m	54.54m
C3							
T1	4	-	-	3	2	2.49m	15.10m
T2	9	1	-	6	1	8.32m	16.30m
T3	31	4	1	13	2	14.29m	45.06m
T4	26	7	2	10	8	11.44m	47.36m
T5	15	9	1	8	7	16.53m	26.40m
Total	85	20	4	40	20	54.47m	96.15m

Table 8.5 Frequency Count of Subject 1's Behavioural Data

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

*total time on task =
total time on task - total time on manual/ menu

C1	Consults manual/menu	E advises	E explains	N asks Q	N asks help	Time on manual/menu	Time on task
T1	5	-	2	5	1	1.29m	30.50m
T2	15	4	1	3	-	4.55m	36.12m
T3	14	6	-	3	2	20.06m	47.01m
T4	28	8	-	8	4	12.01m	48.36m
T5	13	8	-	5	5	9.29m	20.31m
Total	75	26	3	24	12	48.00m	*135.10m
C2							
T1	3	5	-	2	2	1.39m	30.54m
T2	10	5	3	2	5	7.12m	21.09m
T3	22	13	7	7	10	14.39m	50.31m
T4	28	2	1	6	7	19.51m	57.34m
T5	11	6	-	2	4	13.17m	27.05m
Total	74	31	11	19	28	56.38m	130.35m
C3							
T1	3	2	-	3	-	1.33m	36.41m
T2	9	8	-	1	2	7.18m	22.41m
T3	16	7	1	5	3	26.31m	36.09m
T4	9	6	1	6	2	30.06m	49.41m
T5	8	16	2	9	7	9.13m	36.04m
Total	45	39	4	24	14	74.41m	106.35m

Table 8.6 Frequency Count of Subject 2's Behavioural Data

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

*total time on task =
total time on task - total time on manual/menu

	Consults manual/menu	E advises	E explains	N asks Q	N asks help	Time on manual/menu	Time on task
C1							
T1	9	-	-	-	-	6.37m	21.05m
T2	6	-	-	-	-	1.43m	11.24m
T3	14	3	-	-	-	12.17m	25.00m
T4	22	1	-	-	-	12.11m	29.07m
T5	13	1	1	-	-	20.41m	28.09m
Total	64	5	1	0	0	53.29m	*61.08m
C2							
T1	-	1	-	-	-	-	13.26m
T2	5	3	-	1	-	9.09m	15.51m
T3	21	8	2	2	3	11.35m	30.49m
T4	28	4	-	1	-	12.59m	37.53m
T5	19	6	2	2	-	13.40m	27.20m
Total	73	20	4	6	3	47.23m	76.56m
C3							
T1	13	-	-	-	-	14.05m	35.13m
T2	7	-	-	-	-	1.50m	11.31m
T3	16	3	-	-	-	10.33m	29.37m
T4	34	7	1	5	1	30.20m	72.19m
T5	12	7	-	1	2	20.29m	29.38m
Total	82	17	1	6	3	77.17m	101.01m

Table 8.7 Frequency Count of Subject 3's Behavioural Data

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

*total time on task =
total time on task - total time on manual/menu

significant are explicitly stated and all coefficients are included in parentheses.

8.4.1 The Embedded Figures Test (EFT)

The EFT scores appear to correlate negatively with the frequency count for manual consultation for 2 of the 3 word processors, but these correlations are not statistically significant (using WS=-0.967, using WP=-0.916, using PW=+0.900). These results contradict previous findings from Study 1 which demonstrated a positive relationship between the EFT and frequency of map consultation. There may have been stronger factors influencing the number of times subjects consulted the manual, namely, the type of interaction that occurred between expert and novice and overall performance. S2, who showed characteristics of being field independent, tended to look at the manual more than the other two subjects but in addition to this, he showed superior performance in terms of number of moves over the other subjects. S1 showed characteristics of field dependence and tended to look at the manual less than the other subjects but also tended to ask for more help and ask more questions than any other subject. Consequently, she obtained her answers from the expert rather than the manual.

The EFT did not correlate with the number of moves taken to complete the tasks, as expected from the results of Study 1. This was mainly due to the fact that S3 received very little help from the expert, so the difference between her mean number of moves and the others could possibly be due to the fact that she tried alternative solutions before asking for help.

8.4.2 The Matching Familiar Figures Test (MFFT)

Time to respond in the MFFT correlated positively, but not at a statistically significant level, with the number of moves the subject took to complete the tasks (WS=0.984, WP=0.996, PW=0.965) The correlation between the MFFT and number of moves taken using WP is nearing significance at $p<0.05$; that is, the more reflective the subject is, the more moves s/he tends to make. This result is counter-intuitive: one would normally expect impulsives to make more moves as they would be more likely to press keys in a random manner. Alternatively, reflectives might attempt various solutions before asking for help. This is corroborated by the fact that reflectives tend also to be divergent thinkers and S3 was divergent in her thinking strategies. S3 also asked fewer questions and asked for help fewer times overall, than the other two subjects. She was also given less advice and fewer explanations from the "expert", working almost entirely on her own. When many measures of performance are taken, it is possible to see how factors interact to modify the subject's performance.

8.4.3 Divergent Thinking

Divergence and time to complete task for the WS word processor correlated at 0.997, $p<0.05$. Divergent thinkers tend to do better in terms of time on task for the more complex word processors. One possible reason is that they prefer diversity which complex systems can provide. The possibility that subject 3 (divergent) performed well because she received more help can be eliminated. It is quite apparent that she received far less help than the other two (less divergent) subjects. However, she interacted with this particular word processor after using the other two (the order of introduction to the word processors was rotated) so there may have been a learning effect.

8.4.4 Extraversion

Extraversion, as measured by the EPI correlated negatively with asking for help while interacting with the WP word processor at -1.00, ($p<0.01$). This suggests that as extraversion increases, the tendency to ask questions decreases. Again, this seems to be counter-intuitive and suggests that there were stronger factors than extraversion which influenced the rate of asking questions. Alternatively, this result may be spurious since the subjects differed only by one or two points from each other on the introversion-extraversion scale of the EPI.

8.4.5 Instability

Instability, as measured by the EPI is negatively correlated with the number of moves for WS (-0.999, p<0.01), WP (-0.995) and PW (-0.885). Thus, less stable subject used fewer moves. Less stable subjects tend to proceed with caution and may require confirmation before proceeding. This result most likely reflects the amount of help the unstable subject received during the interaction.

8.4.6 Styles of Learning

8.4.6.1 Holistic Orientation

Holistic orientation (redundant holist) correlated positively with the number of moves taken to complete the tasks for WS (+0.934), WP (+0.995) and PW (+0.977), suggesting that as holistic orientation increases so does the number of moves. The redundant holist tends to "globetrot" a great deal and therefore stray from the problem at hand. This may cause the number of moves to increase, in particular, if the redundant holist is also a divergent thinker.

Holistic orientation was also found to correlate with expert advises but only for WS (-1.00). That is, for the more complex word processor, redundant holists receive less advice from the expert.

8.4.6.2 Holist Style

Holist style correlated positively with the frequency with which the subject consulted the manual while interacting with WS (0.924) and WP (0.972) but not for PW. Holistic style also correlated positively and significantly with time spent consulting the manual for WS (1.00, $p<0.01$) and PW (1.00, $p<0.01$), but not for WP (0.894). Overall, comprehension learners tend to spend more time consulting the manual. However, it is also possible that the more help a subject receives, the less time is spent consulting the manual. From the data this appears to be true (see Tables 8.5-8.7).

Holist style tends to be negatively correlated with the frequency with which the expert explains, particularly for the more complex word processors such as WS (-1.00, $p<0.01$). In addition, there appears to be a trend towards a relationship between holist style and the number of questions asked while interacting with WS (-0.982), WP (-0.901) and PW. These correlations, however, are not statistically significant, but suggest that comprehension learners as opposed to operation learners tend to ask fewer questions because they spend more time consulting the manual. Comprehension learners may prefer to learn from the manual and to spend more time understanding basic concepts rather than attempting to do the tasks in vivo.

8.4.6.3 Serialist Style

Serialist style was found to be positively correlated with the time taken to complete the tasks using the WS word processor only (0.997, $p<0.05$). This suggests that comprehension learners tend to perform better using a more difficult word processor. Complex word processors may entail greater understanding of the functions rather than focussing on simple operational details, that is, whether it works or not.

8.4.6.4 Versatility

Versatile learners are supposed to be the most successful learners because they are able to adapt to the style of learning which will be most effective for each problem. However, the only significant correlation found was with E (Expert) advises for WP (+0.999, $p<0.01$) suggesting that the more versatile learner obtains more advice from the Expert.

8.4.7 Number of Moves

The number of moves taken by each subject for each word processor tended to correlate positively (WS and WP=0.965, WS and PW=0.902, WP and PW=0.984). The slightly higher correlations tend to be between word processors that are most similar in difficulty.

For WP, there are associations between number of moves and frequency of advice-giving (-0.930) and time on task (-

0.955) but unfortunately they do not reach statistical significance. Thus, the tendency is that the more advice a subject is given, the fewer moves that subject makes and the more time the subject spends doing the task. One would expect the former inverse relationship as the subject is receiving more clues as to how to proceed. One explanation for the inverse relationship between moves and time to complete the task is that the time that the expert spends giving advice or explaining or similarly the subject asking questions has not been subtracted from the overall time, in which case it is possible that the actual time spent on the task for all subjects is fairly similar.

For WP, the number of moves was most highly correlated with the frequency of asking questions (-0.959), although significant levels were not reached. The greater the number of moves made, the fewer the number of questions asked by the subject. This can be explained by the fewer prompts being given.

For PW, the number of moves was correlated negatively at $p<0.01$ level with asking for help (-1.00), and non-significantly with frequency of asking questions (-0.991) and the expert explains (-0.933). These factors were interacting to reduce the number of moves taken to complete the tasks.

8.4.8 Word Star

Correlational trends between frequency of consulting the manual using WS and frequency count of consulting the manual using WP (0.988) but not between the former and frequency of consulting the manual using PW (-0.758) were found. Given this, the expectation that the number of times the manual was consulted and time spent consulting the manual also correlated was borne out not only within the same interaction (WS and WS) (0.922), but also across word processors: WS and WP (0.997, $p<0.05$) and WS and PW (0.928). There was also an expected inverse relationship between number of times the manual is consulted and frequency of expert explains (-0.930). That is, the more often one consults the manual the less likely one will obtain an explanation from the expert.

Positive relationships also occurred, as expected, between advice-giving and asking for help (0.929), that is, if one asks for help one is likely to be given help. A correlation was also observed between how often the subject asked a question and how often the expert gave an explanation (0.979) and how often the subject asked for help and how often the expert gave an explanation (0.966). Similarly, asking for help and asking questions is highly correlated (0.999, $p<0.01$). Obviously, all these factors are closely interrelated.

Explanations and time spent consulting the manual also produced significant correlations (-1.00, $p<0.01$) not just within the same interaction, but also between explanations using WS and consulting the manual using WP (-0.901) and between explanations using WS and consulting the manual using PW (-1.00, $p<0.01$). This suggests that subjects performed consistently throughout the study.

Time spent consulting the manual while interacting with WS correlated positively with time spent consulting the manual while interacting with PW (1.00, $p<0.01$) but none of the correlations reached significance levels for other combinations.

8.4.9 Word Perfect

A correlation was observed between the frequency of consulting the manual and time spent consulting the manual but did not reach significance (0.974). The frequency of manual consultation correlated better with the time spent completing the tasks (1.00, $p<0.01$), whereas no relationship emerged between time spent consulting the manual and time spent completing the tasks (-0.27). Advice-giving and time on task correlated positively (0.997, $p<0.05$).

Asking questions and asking for help did not correlate (0.400), although asking questions using WP correlated highly with asking questions using PW (1.00, $p<0.01$) and not so highly between WP and WS (0.967). Again, this demonstrates consistency in subjects' performance.

8.4.10 Perfect Writer

Trends were observed between time spent consulting the manual and time spent completing the tasks (0.980) and between asking questions and asking for help (0.989).

8.5 Summary

Despite the small subject size and the high correlational values required in order to reach significance levels, there are patterns in the data that are accountable. For example, if a subject spends most of his/her time consulting the manual, the frequency of asking questions or the Expert advising or explaining is reduced, and time spent on task is likely to increase. Another example, is that the more questions the subject asks, the higher the probability that s/he will be given help and thus reduce the overall number of moves and time on task.

The main focus of the study, however, was to see how the subjects chose to interact with the word processor and how they deal with the problems that they faced.
The origin of the choices were expected to lie in the subject's particular personal style. Many of the correlations have given some indication of how each component or dimension of style affect behaviour and these relationships are substantiated by the correlations and

definitions given in Chapter 7.

In Chapter 7, predictions were made from the sum of the subject's attributes as measured by the personal styles tests. It is now possible to compare predicted performance with actual performance.

Table 8.8: Actual Performance for S1

	No.of Moves	*Consult M	Time on M	#Consult E	%Time on T
PW	1.53	85	54.47	93	96.15
WP	2.80	66	29.45	58	54.54
WS	3.15	35	21.56	162	135.44

*Consult M= Consult Manual
#Consult E= Consult with the "expert". This refers to the total of expert advises, expert explains, novice asks a question and novice asks for help.
%Time on task refers to total time spent completing the task minus the time spent consulting the manual.

Preference: PW

In terms of number of moves, S1 showed a superior performance using PW over WP and WS. However, when the extent of expert-novice interaction is taken into account, S1 performed better using WP. In addition, S1 spent less time on task using WP than when using PW or WS.

In terms of performance, S1 is most suited to WP. In terms of S1's personal style, it was predicted that she would be most suited to PW or WP.

Table 8.9: Actual Performance for S2

	No.of Moves	Consult M	Consult E	Time on M	Time on T
PW	2.39	45	81	74.41	106.35
WP	2.95	74	89	56.38	130.35
WS	3.11	75	65	48.00	135.10

Preference: WP

In terms of number of moves, PW appears to be the easiest with which to interact. When other factors are taken into account, S2 performed better using WS, in terms of the amount of help received during the interaction and also in terms of how much time was spent consulting the manual over PW and WP, but not in terms of time on task. In general, S2 is most suited to WS. This is confirmed by S2's personal style.

Table 8.10: Actual Performance for S3

	No.of Moves	Consult M	Consult E	Time on M	Time on T
PW	3.89	82	28	77.17	101.01
WP	5.39	73	35	42.23	76.56
WS	3.57	64	6	52.37	62.08

Preference: WP

S3's performance on WS was superior in all aspects to WP and PW. S3 is best suited to WS and this is consistent with S3's personal style.

8.6 Discussion

From the performance data presented above, it appears to be generally true that performance can be predicted from a user's cognitive and learning style and personality characteristics. However, we do not yet know which factors are most important in influencing performance. Furthermore, if it is possible to categorise users in order to assign them to the system most appropriate for them, then how feasible is it to employ testing on such a large scale?

If the design objective is to give users, computer systems which are most suited to their individual needs, then level of experience may well be the most economical way of making that decision. But in terms of self-adaptive interfaces, it may be possible to observe how the different elements of personal styles interact and then to construct a table of "dominance". For example, how does a factor modify itself when interacting with another factor? Or which factor takes precedence when there are two or more opposing factors? A table of dominance is attempted in part of the analysis in Chapter 9.

Stereotypes of users in terms of their respective courses (Arts, Sciences) have already been suggested and implemented to a certain extent by Rich (1983). In the same way, it seems possible to construct stereotypes of users' behaviour in terms of cognitive and learning styles and personality factors. These stereotypes would reside in the user's history and could drive the intelligent front end by making decisions as to the "terms" of the interaction. The "terms" of the interaction could be:

- type of communication mode to be used
- whether and when to change the communication mode following practice
- the extent and type of explanation
- the rate of feedback to the user
- the amount of support and advice in terms of help facilities.

The user's history could record the user's performance and any difficulties experienced. This would allow the intelligent front end to modify the "terms" of the interaction. The performance, as predicted by the stereotype, should always be commensurate with the actual performance. If there is a mismatch, then it would also be possible to modify the stereotype.

In summary, a self-adaptive interface founded on the personal style of the user appears to be both feasible and implementable. The result would be, as Rich suggests, a system that appears different to everyone.

One final point relates to the incompatibility between the word processor predicted to be most suitable for the individual users and the word processor that each preferred to use. From the verbal protocols, it was apparent that none of the subjects enjoyed using WS. S2 and S3 both preferred WP while S1 preferred to use PW. However, her main reason for not liking WP was primarily due to her dislike of the manual. The documentation supplied did not contain detailed instructions. The amount of documentation was minimised in order to avoid supplying the subjects with a complete model of the word processors.

8.7 Conclusions

An attempt has been made to predict subjects' performance from their personal style. One difficulty is that there exists many types of information in an interaction which will all contribute differently to performance. The empirical problem is how to isolate the different types of behaviour and demonstrate that certain traits have a differential effect on behaviour.

A small n design was employed because users of computers differ on many dimensions, therefore making intrasubject differences important. Further research is needed to continue investigations of methodological procedures. For example, is it feasible to use this type of methodology on larger n sample? A follow up to the study reported here should attempt to categorise users according to their personal style and then give them matching or mismatching materials (see Pask and Scott, 1972).

Chapter 9 Verbal Protocols

9.1 Introduction

The usefulness of verbal protocols has been questioned by some researchers (e.g., Nisbett and Wilson, 1977). They found that retrospective verbalisations about subjects' motives for their behaviour were no more accurate than the observers' judgements about those motives. Others such as Ericsson and Simon (1980, 1985) have created a model of how different types of verbalisations affect the task and influence the underlying structure of cognitive processes.

Verbalisations can be divided into concurrent and retrospective, that is, verbalisation during the task and verbalisation after the task is completed, respectively. Ericsson and Simon's model assumes that verbalisations are of 3 types: level 1 refers to a direct articulation of information, accessible as a verbal code. Level 2 verbalisations refer to a recoding of non-propositional (non-verbal) information into a verbal code without any additional processing, and level 3 verbalisations are articulations after the information have been subjected to the person's filtering and other inferential processes which have modified the information.

Ericsson and Simon predict that level 1 verbalisations do not affect cognitive processes because the information is easily accessible as a verbal code. If the information is non-propositional, then task performance may be slowed down and the verbalisation may be incomplete. However, the underlying cognitive processes will change minimally.

The use of verbal protocols is not extensive in human computer interaction. Those who have used the technique for exploring how learners learn to use computers (e.g., Mack et al, 1983, Carroll and Mack, 1984) have relied solely on qualitative descriptions. Mack et al (1983) requested novice computer users to verbalise while learning to use a word processor and demonstrated their misunderstandings with interesting excerpts from their reports.

Mack et al (1983) classified the problems that their subjects encountered into eight main categories. These will be briefly summarised.

i. Learning is difficult: Learners were found to blame themselves and think that they were stupid. Learners also took longer than expected to learn basic editing skills. In particular, they had problems with applying the knowledge they had acquired in the training session to the test session.

ii. Learners lack basic knowledge: Learners did not possess sufficient general information about how computers work and sufficient specific information about how the text editor worked. Without this knowledge it is difficult for learners to make intelligent guesses as to what has gone wrong. In addition, learners had difficulty understanding basic ideas and technical jargon. Again this leads to difficulty in identifying what is relevant to solving the problem which results in the user making incorrect assumptions between cause and effect.

iii. Learners make ad-hoc assumptions: despite the learner's lack of knowledge, s/he will attempt to make sense of what is going on. Often these interpretations are fallacious and absurd. Learners need more knowledge to help them construct accurate interpretations.

iv. Learners generalise from what they know: Analogies are useful to the learner but there are always differences between the analogy and the object e.g. typewriter and word processor. In addition, learners also use knowledge that they have just acquired in the interaction to help them. The problem is of maintaining consistency in procedures and this is exacerbated by the fact that designers and users (in particular novices) differ in how they view operations.

v. Learners have trouble following directions: Learners often believe they are better "doing" rather than reading the manual. They also have much information with which to contend such as reading the manual, pressing the correct keys, following the responses on the screen and thinking about what is going on. This may lead to the learner losing track of where s/he is and may also not be conducive to effective learning. Mack et al suggest that one way to overcome this problem is to give learners more control over their training activities.

vi. Problems interact: It is difficult for the learner to identify the cause of the problem because so many things are happening.

vii. Interface features are not obvious: Some aspects of word processing appear arbitrary such as pressing a special key before typing in. These types of commands have to be learned as a requirement. Learners would encounter fewer problems if more of the interface features were obvious. In addition, poor feedback messages are also problematic.

viii. Help facilities do not always help: One reason for this is that the learner does not possess the jargon to understand the help facility. Another reason is because learners do not always know what the problem is and how to describe it.

In the study described in Chapter 8 the sessions were less constrained than those of Mack et al. Subjects were asked to verbalise but the sessions were directed more towards an "interactive" dialogue with the "expert". One of the reasons for this arose from the need to look at the frequency of asking for help and support (see behavioural analyses in Chapter 7). Second, it was believed that richer verbal protocols would emerge if the dialogues were interactive. Some sessions also contained a large amount of prompting depending on how well the subject was able to verbalise. The prompts always involved the subjects' current move or problem, rather than prompts for less relevant information, which may have substantial effects on performance (Ericsson and Simon, 1980). The differences in methodology have resulted in differences in classification. While Mack et al's classification tends to be general, many of the problems encountered in Study 4 are subsumed in their classification. Their comments and examples of dialogue were familiar and echoed many of the problems and feelings of my subjects. However, it was felt that the problems could be subjected to a finer grained classification.

The utilisation of verbal protocols in this study is aimed first, at producing a classification of types of misunderstandings and illustrations through sample extracts and, second, to include a quantitative analysis of the data to provide insight into which misunderstandings are typical of particular word processors. More importantly, a quantitative analysis can give insight into the similarities and differences in the types of problems the subjects experience and an indication of whether these similarities and differences can be accounted for by the subjects' cognitive and learning styles and personality characteristics.

Misunderstandings were classified according to the dimensions set out below. The meaning of those dimensions which are not explicit to the reader are clarified.

i. Basic editing problems - refers to problems of insertion and deletion.

ii. Incorrect command used - refers to difficulties arising as a result of N using the wrong command, including accidental use of a wrong command.

iii. Locating a command - refers to N being unable to locate a "known" command either in the manual or on the screen.

iv. Did not read screen - refers to difficulties arising as a result of the N not reading or misreading the screen.

v. Does not know how to carry out a command.

vi. Difficulty moving through the menu system or getting "lost" - refers to obtaining an incorrect menu or not knowing where one is in the menu.

vii. Incorrect reasoning or incorrect association between command and consequence.

viii. Don't understand the function of a command - refers to problems arising because N does not understand what a particular command does.

ix. Misunderstanding the function of a command - refers to difficulties arising through N confusing the function of one command with another command.

x. Wrong cursor position - refers to errors due to the cursor being positioned incorrectly.

xi. Changes not understood - refer to those changes having occurred to the text but their meaning being unclear to N, e.g. "underlining" in Word Perfect is shown in the colour green.

xii. Misreading the manual - refer to those difficulties arising because N has not read the manual or followed the instructions.

xiii. False assumption that a mistake has occurred - refers to problems arising because N believes incorrectly that s/he has made an error.

xiv. Typing in at beginning - refers to the problems encountered while starting to type or typing in the text.

xv. Typeover versus deletion - refers to N becoming confused as to whether to typeover or delete and type again.

xvi. Inconsistent procedures - refers to N experiencing difficulties performing a command despite being familiar with the "procedure" of the command. For example, Word Perfect uses the same procedure for "bold", "underline", "centre" and "flush right".

xvii. Does not realise a change has occurred - refers to N being unaware that the word processor has changed the contents of the screen.

xviii. Incorrect spelling/case - refers to difficulties experienced because N has spelt a word incorrectly or used the wrong case.

9.2 Qualitative Analyses

Days 1 and 2

Examples of verbal protocols illustrating some of the misunderstandings experienced by subjects for days 1 and 2 are given below. These do not constitute entire transcripts but excerpts which are interesting and serve to highlight the difficulties and misunderstandings experienced by typical novice users.

Basic editing

Example 1:
N experiences problems with basic editing commands. N tries to "push" the extra text off to the next line but this does not work without a hard return having previously been inserted or the margins reset.

N: If I used ^T

E: What do you think ^T would do?

N: Move the word right, hopefully. I don't know. All I'm trying to do is push it off the end to the next (line). Whether it'll do that I don't know. No, ^T is from the delete thing, of course. I don't know.

E: You've no idea at all? You can't think of anything you used for the other word processors which might help?

N: I'm sure there's something but I can't remember.

Example 2:
N has problems inserting lines. One of the problems is that N sees her problem as moving the already existing text down one line rather than inserting a line space at the top.

E: What are you trying to do?

N: Get "Lone Pine National Park" down one line.

N presses the cursor key a number of times and then looks in the manual.

E: What are you looking for? To move the line down?

N: Yes.

N moves the cursor to the end of the line.

E: What are you trying to do?

N: I'm going to type in the date.

E: But it's on the wrong line for the date.

Some of N's procedures are rather complex!

N: Well, I was going to type the date, split the line and push it up.

E: Oh, o.k.

.
.
.

N: Well, that's no use

N has pressed the return key and the date has gone down to the next line.

E: What are you looking for?

N: Anyway to move this damn

E: So you've gone back to trying to move the line down?

N: Yeah.

N presses the return key.

E: What made you think of that?

N: Why did I think of that Well, it's one way of moving the sentence down.

Incorrect command used

Example 1:
N is unable to find the underline command so attempts to insert the command in "reveal codes" which does not permit codes to be entered in this way. This is an example of finding the next best way to achieve a goal.

N is looking up "reveal codes".

E: Ah, you've used the Alt F2 (reveal codes key). What's that? Is that what you wanted?

N: What!...

E: What does it tell you?

N: I think it's a breakdown of what I've got - centering, that's right... so I just have to put in an extra command where my cursor is.

E: What did you press there?

N: Shift U.

E: Ah, I see, you've put in shift U to tell in to underline it. Is that what it tells you to do there (in manual). Do you think that's right that you can enter your own codes? How did those other codes get there?

N: They're obviously automatically clocked up as they're used.

E: So do you think you should be able to go in and put your own codes in?

N: Well, it might help as I can't see any other way of underlining it.

E: So what have you been using to do all the other things?

N: Yes, I know but I can't find any other way.

Example 2:
N knows that "justify" tidies up the text but does not realise this command cannot be used with single lines of text.

E: What are you doing now?

N: I'm tidying up.

E: What are you tidying up?

N: The state of the thing. It probably won't do much.

N uses justify to clean up the title.

N: Well it's done nothing.

<u>Locating a command</u>
Example 1:
E: Where else do you think might be a good place to look?

N does not know where to look for the commands.

N: Retrieve?

E: No.

N: Ah, it's under "exit".

E: It's under exit?

N: Yes because it then asks "replace filename". "Type in Y to replace the file on this with the one on the screen". That's all you do, isn't it?... No it isn't it, is it?

Example 2:
It appears that novices have problems locating commands. Their idea of where a command should be found are not always the same as those which have been chosen by the designer.

N: I'm not sure exactly where I should be looking. I expect it should be layout. I can't see it there but I'd imagine it's there.

E: You think it's in layout?

N: Yeah, I would have thought it's that sort of command.

E: Are you going through the manual indiscriminately?

N: Well, I'm not going through "revising a document", but I'm going through it chapter by chapter...

Did not read screen

Example 1:

N: How do I insert a character now?

E: Have a look at the summary. Are you not looking at all at the menu?

N: No, I wasn't doing. Should I?

.

.

.

Can I delete a whole word? (consults summary and does not look at the menu).

Although the menu on the screen shows all the deletion, insertion and cursor movement commands, there is far too much information on the screen. N is totally oblivious to this information.

Example 2:

Sub commands (under function keys) in Word Perfect appear at the bottom of the screen. Novices often do not see this and when they do, they take very little notice of it when they are following instructions in the manual. Misreading of instructions often lead to problems which might not exist if the Novice had been reading the screen.

N: It doesn't seem to have done anything. It's not done what I expected... I seem to have got a "4" at the top of the document. I'll get rid of that. "Shift F8 comma 4". That's what I did, but it didn't work. I've obviously done something wrong.

E: What did you expect to see?

N: That (points to a description in manual) in the brackets.

E: You didn't see line spacing?

N: No, I'll try it again and see if I've missed something or pressed the wrong key. Shift F8 comma 4. Return that. It's done exactly the same thing again. Why is it doing that?... I don't think I had to do the comma.

Difficulty moving through the menus system

Example 1:

N: Do I just want to press B to start blocking? (N presses the downward cursor key which corresponds to ^X which exits N from Word Star as N is already in the block menu so X has a different meaning)

.
.
.

It's started a new file.

E: Why do you think that is? So that's not the file so what are you going to do?

N: I don't know (N presses ^J). Well, that doesn't help, does it?

E: That doesn't help?

N: No, the help menu doesn't help.

E: Why?

N: Because it doesn't tell me how to get into a file. "Index of commands" (N presses I) What do I do now?

E: Do you want to have a look at your summary.

N: Which one of them gets you in, if it's not edit, which I did?

.
.
.

E: What are your possible moves now?

N: What do you mean?

E: You know you don't want a new file so how do you get out of this?

N: I don't know what my file's called.

E: It's called XXXX.

N: What do I want to do now? How do I get out of this? "End edit, save file". Thought I was supposed to get into K (menu) then why am I in J (menu). Am I out now?

E: You're out of the help menu.

N: But I'm not out of editing, am I?

E: No. Why do you think that when you pressed ^K you got some more of the help menu?

N: Because you always go back to the help menu, don't you?

E: No.

N has not read or assimilated the information from the screen. N presses control keys to get into the menus when she has not yet left the previous menu, nor returned to the main menu. All menus are accessed from the main menu. N becomes confused and does not know whether she has left the menu or not.

Example 2:
Again, N is having problems of "knowing" where she is in the menus. She attempts to use a command from the main menu while in the block menu. The same command from the two menus have very different meanings.

N presses ^C and receives the error message "Error 6: Block beginning not marked".

E: Why did you press ^C?

N: Well, I had to move the cursor down one screen so I can get to the bit I've got to move up.

E: Do you know what's gone wrong there?

N: I think I should have gone down before I got into K (menu).

E: Gone down where?

N: Gone down to the bit I want to move up.

.

.

.

I still don't know how to get it to the bit I want to move it to. I'll get the help menu.

(N is in the K menu and presses J which N believes to be the help menu but means "delete" while in the K menu).

The computer responds with "Name of File to delete?"

E: What've you done?

N: Pressed J.

E: You pressed J on its own?

<u>Incorrect reasoning</u>
Example 1:
Word Perfect does not show justification on the screen. The failure to detect any changes to the text leads N to believe that his last command has justified the <u>left</u> margin, which of course, has always been straight.

N: Has that done it? Has it done it?

E: Does that look justified to you?

N: It doesn't look any different, I'm sure. I'd have imagined it just gets everything together in one line. It doesn't appear to have done it. Has it done it? I'll try again to see if I've missed anything. Ah, so that's just lined up this side.

E: Which side?

N: The left one of the thing, I should imagine. I think.

E: So you think you've justified the left hand side?

N: Yes, but it's a bit silly because it's all... yes because if I've done that it's probably messed everything up. That's exactly the same as before (looks at document). That's the same. If I moved the cursor over to the other side (right side).

N incorrectly believes that he must place the cursor to the right hand side before entering the command.

E: What are you trying to do?

N: I'm trying to justify it from that side. See if it makes any difference. I'm not sure if I'm doing this right. "Cursor to where you want to start the justification" Do we want it right at the end of the line or smaller? (i.e. further in). Let's try it here. That's had no effect whatsoever. I would imagine if I moved it further back along the line, then it would justify in the right place. The problem is I'm not sure exactly what is required here.

E: You know what justification is?

N: Yes, it's squaring up the text.

Example 2:
Word Star does not disregard "case" and must be instructed to do so when "replacing" words. N correctly inserts the global replace option but types the word in lower case. An identical match is found with only one of the key words. Because one word has been replaced, N incorrectly assumes that the error must be due to misunderstanding the meaning of the "Replace, yes/no?"

N: I think it's probably gone through the entire file and replaced them all... It hasn't done anything.

E: Why do you think that is?

N: I don't know. I think I did it right according to the way I understand the characters. It's only replaced one of them. That's not right. I did something wrong along the line. The only thing I can think of is the yes/no response thing, because up until that point it only asked me what I wanted to do. So I obviously misunderstood that command...

I think I must have chosen the wrong option.

E: So you should have said "no".

N: That's the only thing I can think of because until then everything was obvious. So I'll try it again. I'll try the opposite.

.

.

.

I don't understand why it's done that. "Not found Needles". Why hasn't it found it? Unless it replaced them all before but they were all on the screen so they couldn't have.

E: So what do you think you should do now?

.

.

.

N: It changed one but it didn't change the others. I don't know really understand why. Unless it's only searching through a certain amount of text at a time. But that would seem silly. I would imagine that the global replace thing is what it sounds like...

<u>Don't understand the function of a command</u>

Example 1:

Terminology that N is not familiar with lead to problems in selecting the appropriate commands. In addition, N must consider "dot commands" which are typed in rather than selected from a menu as the majority of commands.

N: What does "standard pitch" mean?

E: You don't need that.

N: But it says the command is "PN" so how do I do that? Don't I have to go into print? (menu) What are "dot commands"? You've not explained to me what a dot command is.

E: What are you going to do about it?

N: I expect I'll go into the help menu and it'll be of no help whatsoever... So can I just put .PN at the top of the file?

<u>Misunderstanding the function of a command</u>

Example 1:

N sees no difference between the functions of "save" and "rename".

N: Rename, why can't I just save it with a different name?

E: Then you'd have 2 copies, wouldn't you?

N: There isn't a "rename". If I just save it and call it a different name, would that still save the original file?

E: That would save the original file. There is a "rename".

Example 2:
N understands the term "justify" but cannot distinguish between the two commands "to justify" and "the justified mode".

E: Is that the one you want?

N: I think so. Justify and justified mode.

E: What's the difference between justify and the justified mode?

N: I don't know to be honest. Um, as far as I can see from this (justify mode) justifies the whole document, I think.

E: So that's what you're going to use?

N: Yes, I think so...

E: Is that justified?

N: No, I don't think so. It hasn't done anything, has it? Ah, yes, it is. It says "justify" at the bottom.

E: You're in justified mode but is your text justified?

N: No, well, it doesn't look like it.

<u>Wrong cursor position</u>
Example 1:
N splits the line by pressing "return". Instead of pressing the corresponding delete key to join the lines again...

N: I pressed, well, it's just moved the rest of the line down from where the cursor was. What do I do now? This is ridiculous. I think the easiest way. Is that "Christ" centered? It is. It's just centered in the middle by itself, isn't it?

E: Mmm.

N: I think the easiest way is to delete that line and retype it in the centre.

E: You're going to do that, are you?

N: I'm sure there's an easier way. There's probably a way of getting it back, but I think that's the easiest way.

Changes not understood

Example 1:

Previously, N has experienced problems with cursor position before entering a command. N similarly believes that the result of this command is due to incorrect cursor position.

E: What's happened there?

N: Did I need the cursor at the beginning? "Remove underline from a marked area".

The "bright lights" show underlining, but clearly N does not agree!

E: Is that underlined at the moment?

N: I don't know what it is to be honest, I just pressed the cursor and it's done something wrong. What's that done? Why is Christmas in bright lights?

E: I don't know, why is it?

.
.
.

Do you think "Christmas" is underlined?

N: It looks bold to me. Is it bold?

E: No

.
.
.

N: Does that count as a marked area?

E: Does what count as a marked area?

N: That in bright letters.

Misreading the manual

Example 1:

Incorrect and misreading of documentation appears to be common.

N: Because that's what it says there in the manual. That's a "y", is it not?

E: Yes, but did you press Alt F6?

N: No.

Example 2:

N: Hard return?... Does that mean search key in the F thing?

E: What do you think? You don't know?

N: No.

E: Have you done it?

N: It said "please wait".

E: Did it?

N: I don't think it's doing it (N tries again)

E: What's happened there?

Again, N has not read the manual correctly. N knows that hard returns enter commands, but with the "replace" command, the sequence is search and then replace thus the Alt F2 key must be pressed a second time to obtain "replace".

N: I don't know (N tries again). What's it doing this for? (N hits the key randomly).

E: What's bothering you?

N: It keeps coming up with hard return signs.

<u>False assumption that a mistake has occurred</u>
Example 1:
Extra line spaces have to be inserted before they are taken to be additional spaces (in this case at the top of the document) required by N. Many word processors do not show extra line spaces at the top of a document unless the N has inserted them. If N "pages up" the document, it will normally stop at line 1. But for Word Star, the document does show extra line spacing not inserted by N but concurrently reads the first line of text as line 1. N has problems understanding that these extra lines at the top of the document do not in reality exist.

E: You don't think you should use "screen up".

N: Um, I'm sure I deleted it (line spaces) before. But it's obviously not that. I can't remember.

E: Do you think it might be something to do with where you have your cursor?

N: I would imagine so. I can move the cursor up and down but it doesn't do anything, because it still moves back again (leaving the additional spaces at the top of the document). I can't remember.

E: Why don't you try using "screen up"?

N: "Screen up" what's that? Oh that, ^C? (N does it) It hasn't done anything.

E: What are you thinking?

N: It looks to me as if I've got the same space there, so that hasn't done anything, has it? All that it's done is to move the cursor up and down the screen. What you want me to do is get rid of the spaces at the top of the file, don't you?

E: How do you know that is space at the top of the file?

N: Because the cursor's moved up to the top?

E: So if I used ^W, is that not more space at the top (E demonstrates).

N: Well, I would have thought it was, but it's not is it? I see what you mean now. It's not, is it? or is it? I would have expected that was the amount of space at the top of the document.

.
.
.

E: So you've got to distinguish between what's on the screen and what's on the page.

Typing in at the beginning

Example 1:

E: I noticed one thing you did on the other two (word processors) is you just let the sentences wrap around but on this one you decided to press return after each sentence. Do you know what I mean?

N: Yeah. Well, I did wonder about it at one point, whether the screen was deliberately big (wide) to be able to accommodate something very large across the page versus something which was going to be justified or what size of paper they were going to use to print.

E: So how far did you think it was going to go over?

N: I think right up til the end of these (points to return signs at far right hand side of the screen).

E: So that's why you put in the returns?

N: Yes, otherwise they would have all been out of synch and these wouldn't have been paragraphs.

Inconsistent procedures

Example 1:

Both centre text and flush right are followed by the downward arrow key to finish the command. Without the downward arrow key, the text remains to the far right hand side. Alternatively if the text is blocked, the downward arrow key is not used.

After having experienced the same problem with centering text...

E: What's happened there?
N: It's gone far too far to the right (N presses the return key). I'll try again. What's a soft return and a hard return? I suppose...

E: What are you thinking?

N: Well, you can start the flush right then type it (the date) in, can't you? I should imagine it should be easier.

E: Do you not understand what it says?

N: I don't know what a soft and hard return is.

<u>Does not realise a change has occurred</u>
Example 1:
(N sets the right margin to 60)

N: Did that do anything?

E: Do you think it did?

N: It doesn't look like it did... where will it show you it's set it anyway? Oh, do I have to put in "COL 65"?

N does not realise that the ruler shows the settings, and does not notice any change having occurred which leads her to believe that the command has not been executed. Next, N incorrectly believes COL60 should be typed in. The ruler disappears because COL60 is too long and the right margin has been set at 0.

E: What's happened there?

N: Nothing. What happens if I press escape? (The ruler disappears).

E: What's happened to your ruler?

N: Oh, why did it do that? Do I have to get that back?

E: I think you'd better try and get that back.

N: Could you tell me how please?

E: No.

(N looks in manual and then uses help menu)

These are just some of the examples of the transcriptions made from the video recordings. They help to illustrate some of the misunderstandings experienced by fairly naive computer users. The following section examines further these misunderstandings in a quantitative manner and a classification of the misunderstandings is attempted.

9.3 Quantitative Analyses

The quantitative analysis of the data consisted of conducting a frequency count for each misunderstanding revealed in the verbal protocols through the qualitative analysis. The main reason for conducting a quantitative analysis was that it was thought that it might give some idea as to the similarities and differences between subjects in the types of misunderstanding encountered. Second, the quantitative data analysis might demonstrate whether a certain type of problem clusters for a particular word processing package. No inter-rater reliability was established since the data were quantitative.

9.3.1 Days 1 and 2

All day 1 and 2 video tapes were viewed and each misunderstanding classified. A sampling technique was not used because it was felt that this would bias the analysis as it cannot be assumed that the incidence of misunderstanding is randomly distributed in time. Misunderstandings were classified according to the dimensions set out earlier in the chapter.

9.4 Results

The types of difficulties experienced were rank ordered for all the subjects' data for all word processors, and similarly for each individual word processor. Table 9.1 depicts the frequency of misunderstandings for all word

	Total	C1	C2	C3
Basic editing problems	56 (1)	23 (2)	11 (1=)	22 (1)
Incorrect command used	36 (2)	24 (1)	2 (12=)	10 (5)
Locating a command	32 (3)	12 (4=)	8 (4=)	12 (3=)
Did not read screen	30 (4)	18 (3)	7 (6)	5 (10)
How to carry out a command	28 (5)	5 (10=)	11 (1=)	12 (3=)
Difficulty with menus	26 (6)	12 (4=)	0 (17=)	14 (2)
Incorrect reasoning	24 (7)	9 (8)	9 (3)	6 (9)
Don't know function of a command	23 (8=)	12 (4=)	4 (8=)	7 (7=)
Misunderstands function	23 (8=)	11 (7)	5 (7)	7 (7=)
Wrong cursor position	20 (10)	8 (9)	4 (8=)	8 (6)
Changes not understood	13 (11)	3 (4=)	2 (12=)	2 (13=)
Misreading manual	12 (12)	2 (17=)	8 (4=)	2 (13=)
False assumption of error	10 (13)	5 (10=)	2 (12=)	3 (12)
Typing in at beginning	9 (14)	3 (14=)	2 (12=)	4 (11)
Typeover v's deletion	9 (14=)	4 (12=)	4 (8=)	1 (17)
Inconsistent procedures	7 (16)	2 (17=)	3 (11)	2 (13=)
Does not realise a change	6 (17)	4 (12=)	2 (12=)	0 (18)
Incorrect spelling/case	5 (18)	3 (14=)	0 (17=)	2 (13=)

Table 9.1 Frequency Count of Misunderstandings for all Subjects

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

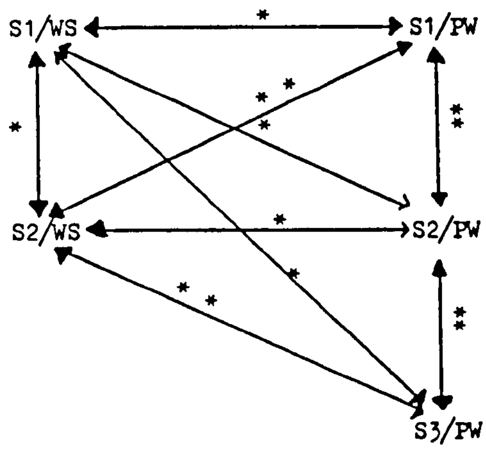
processors in column 1, and then for each word processor in columns 2 to 4. Their rank orders are in parentheses. These data demonstrate which difficulties are more pertinent to each word processor.

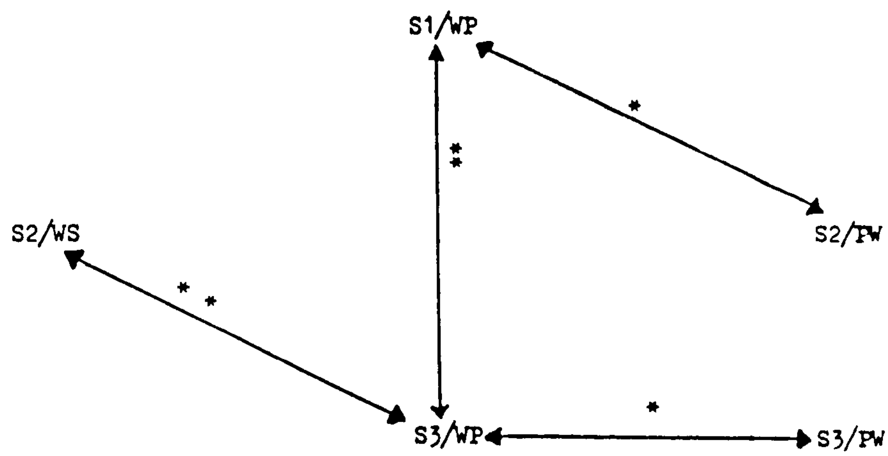
The three most common misunderstandings were: basic editing problems, incorrect command used and locating a command. Subjects had many problems inserting and deleting: these included placing the cursor under the wrong letter and therefore inserting in the wrong place or deleting the wrong letter. Subjects demonstrated little difficulty inserting or deleting text in comparison to inserting and deleting blank lines. Problems typically occurred when attempting to insert a blank line at the top of the page such as the date. This demonstrates the difficulty subjects experienced in using analogies and transferring information learned in different environments.

Incorrect use of a command mainly occurred because the subjects lacked the appropriate command terminology. This problem is linked to problems of command location. That is, if the subject could not locate the command, then s/he often wondered whether the command s/he was seeking was in fact, the correct term for that command.

Fewer misunderstandings occurred for WP (84) than for PW (119) and WS (160). From Table 9.1, it can be seen that the frequency of the problems are distributed differently across word processors. A Spearman's rho correlation was

Fig. 9.1 Correlations between Subjects' frequency of misunderstandings for word processors





*=p<0.05
**=p<0.01

S1=Subject 1
S2=Subject 2
S3=Subject 3
WS=Word Star
WP=Word Perfect
PW=Perfect Writer

computed on the individual subjects' frequency of misunderstandings for each word processor (see Appendix 9.1-9.3). The correlations (see Fig. 9.1) demonstrate that WS and PW were most similar in terms of subjects' misunderstandings and WS and WP were least similar. There are 6 significant positive correlations between WS and PW out of a possible 9 combinations. There are 4 significant positive correlations between PW and WP and only 1 positive correlation between WS and WP. These data suggest that WS and PW are most similar in terms of how they operate, and there are some similarities between PW and WP.

9.4.1 Individual Subject Data

Frequency counts of the misunderstandings for the sum of the three word processing packages were obtained for each subject. Table 9.2 shows the scores rank ordered. This permits a comparison of each misunderstanding for each subject in terms of their position within the set of misunderstandings.

The ranks for each misunderstanding were compared for the three subjects, producing the following seven ties:

Basic editing problems	S1 & S3
Incorrect command used	S1 & S2
Did not read screen	S2 & S3
Wrong cursor position	S1 & S2
False assumption of error	S2 & S3
Typing in at beginning	S2 & S3
Typeover versus deletion	S1 & S2

S1 & S2 = 3 ties
S2 & S3 = 3 ties
S1 & S3 = 1 tie

The following is a comparison of ranks for each subject when the ranks differ by 1 or 2 ranks:

	S1	S2	S3
Basic editing problems	25 (1)	8 (2=)	23 (1)
Incorrect command used	17 (2)	8 (2=)	11 (3=)
Locating a command	9 (9)	12 (1)	11 (3=)
Did not read screen	15 (3)	6 (6=)	9 (6)
How to carry out a command	11 (6=)	5 (8=)	12 (2)
Difficulty with menus	14 (4)	6 (6=)	6 (8)
Incorrect reasoning	12 (5)	8 (2=)	4 (10)
Don't know function of a command	11 (6=)	2 (16=)	10 (5)
Misunderstands function	10 (8)	8 (2=)	5 (9)
Wrong cursor position	8 (10)	4 (10)	8 (7)
Changes not understood	2 (17)	3 (11=)	2 (14=)
Misreading manual	5 (12=)	5 (8=)	2 (14=)
False assumption of error	4 (14)	3 (11=)	3 (11=)
Typing in at beginning	5 (12=)	2 (14=)	2 (14=)
Typeover v's deletion	6 (11)	3 (11=)	0 (18)
Inconsistent procedures	3 (15=)	1 (17=)	3 (11=)
Does not realise a change	3 (15=)	2 (14=)	1 (17)
Incorrect spelling/case	1 (18)	1 (17=)	3 (11=)
Total	161	87	115

Key: S1 = Subject 1
S2 = Subject 2
S3 = Subject 3

Table 9.2 Frequency Count of Each Subjects Misunderstandings for all Word Processors

Locating a command	S2 & S3
How to carry out a command	S1 & S2
Difficulty with menus	S1 & S2
Don't know function of a command	S1 & S2 and S2 & S3
Wrong cursor position	S1 & S3
Misreading the manual	S1 & S3
Inconsistent procedures	S2 & S3
Does not realise a change	S1 & S2
Incorrect spelling/case	S1 & S2

S1 & S2 = 4 ties (+/- 1-2 ranks)
S2 & S3 = 3 ties
S1 & S3 = 3 ties

Others:

Incorrect reasoning	S1 & S2
changes not understood	S1 & S3 and S2 & S3

Total
S1 & S2 = 8 ties
S2 & S3 = 7 ties
S1 & S3 = 5 ties

The above suggests that S1 and S2 are most similar in the way they approach problem-solving tasks, but S2 and S3 also have some factors in common. S1 and S3 differ most in their problem-solving approach. If this is true, that is, that S1 and S2 perform in similar ways, then one would expect them to possess similar personal styles. In fact, S1 and S2 share four out of a possible 6 cognitive and learning styles and personality factors, while S1 and S3 share just two factors and S2 and S3 also have two factors in common.

9.4.2 Misunderstandings and Personal Style

The information gathered previously on each subject's personal style was used to differentiate between the subject's performance in terms of the misunderstandings encountered. By observing the similarities and differences in the subjects' cognitive and learning style and personality, it was possible to cross reference this information with the similarities and differences between the subjects in terms of the misunderstandings. It was thought that this might suggest which characteristic(s) of a person result(s) in a particular difficulty in learning to use the word processor.

The cross reference consisted of first, making comparisons of the rank scores for each misunderstanding across subjects. Ranks between any two subjects which differed by no more than 2 ranks were coded under each dimension subsumed by the generic term personal style which both subjects possessed.

To provide a finer grained analysis it was also necessary to examine large differences in the ranks. The data were further explored by examining subjects' rank scores which were dissimilar by a minimum of 6 ranks. Comparisons for the subjects were made and coded under the cognitive and learning style or personality characteristic which each subject alone possessed. Further eliminations were made by comparing all three subjects ranks together. For example, if S1 and S2 score similarly for wrong cursor position and S3 scores dissimilarly, then the personal style linked to this misunderstanding must be possessed by both S1 and S2 but not by S3. On the other hand, if S3 had not scored similarly or dissimilarly from the other 2 subjects then the misunderstanding cannot be reliably coded.

Reflectivity-impulsivity:
How to carry out a command
Incorrect reasoning
Typeover versus deletion
Inconsistent procedures
Incorrect spelling/case

Field dependence-independence:
Don't know the function of a command
Misunderstands a function
Misreading the manual

Divergence thinking:
Same as reflectivity-impulsivity

Comprehension-operation:
Locating a command

Introversion-extraversion:
Basic editing problems
Incorrect command used
Typing at the beginning

Stable-unstable:
Same as reflectivity-impulsivity

By cross referencing the similarities and dissimilarities of subjects' rank scores for misunderstandings with their similarities and differences in personal style, the number of misunderstandings that may be associated with the subjects' attributes decreased to 12 possible misunderstandings (as shown above).

Unfortunately, it is not possible to tell how much each cognitive style, learning style or personality factor contributed to each misunderstanding. In addition, it is only possible to infer how two factors may work together to alter a user's overall performance.

Field dependence-independence, comprehension-operation: The data suggest that field independents are prone to misreading the manual and this may lead these persons to misunderstand the function of a command. Field independents who are also comprehension learners tend also to experience problems locating commands. This may also be due to skipping over basic details in the manual. On the other hand, field dependents rarely misread the manual and therefore do not often confuse commands. Field dependents who are also operation learners do not have difficulty locating commands: these are the sorts of operations that they learn quickly from the onstart. Field dependent comprehension learners appear to have problems locating commands because, although they do not misread the manual, they tend to focus on global aspects rather than operational detail.

Reflectivity-impulsivity, divergent thinking and stable-unstable: Impulsives are more prone to incorrect reasoning than reflectives. As would be expected, reflectives spend more time contemplating before making decisions or moves. They tend to reason more coherently about their actions. Divergent thinking also appears to be related to ability to reason: those persons displaying more divergence tend to have better reasoning power. But how should reasoning power be defined? It may be that divergent thinkers reason better in unusual or unfamiliar situations while less divergent thinkers reason more logically when there are constraints or the problem has one sole answer. In this type of learning situation (i.e., human computer interaction) divergence may well be the important factor influencing one's ability to reason about one's actions.

Impulsives also tend to be unstable and this may cause fluctuations or inconsistencies in their performances, which may result in incorrect reasoning.

Impulsives seem to be better at carrying out commands but only if they are successful. This is because they spend less time thinking about how to carry out the command and more time actually performing it. This may be particularly true if the impulsive is also an operation learner.

Some commands, such as using typeover or deleting and typing in again, are not easily understood by the users. With such problems, subjects tended accidentally to change typing in modes and then become extremely confused because the word processor was behaving in an unexpected way. Impulsivity, little divergent thinking and instability appear to affect the rate of occurrence of this misunderstanding.

The frequency of inconsistent procedures tends to be greater for reflective, divergent, stable persons. This appears to be counter-intuitive in that one would expect reflective and stable persons to be consistent in command execution. On the other hand, divergence in conjunction with stability may result in the user attempting different ways of solving a problem. Stability permits the individual to try new techniques without fearing the consequences. Little divergent thinking in conjunction with instability would most probably lead users to take the easiest and surest route to success. In addition, if the user is also impulsive, s/he may not pause to think of alternatives but use procedures with which s/he is already familiar.

Similarly, a user displaying little divergent thinking may be cautious and careful over spelling and using the correct case, as will instability in a user.

Extraversion-introversion

As a result of all the subjects scoring high on extraversion, it is not possible to compare behaviour in terms of extraversion/ introversion. All subjects had problems with basic editing and using incorrect commands, and relatively few problems with typing in at the beginning. This may be entirely unrelated to extraversion-introversion; rather, all users may experience these problems to a lesser or greater extent. For example, one of the main reasons for using an incorrect command was because the subject did not understand the terminology used by the word processor. On the other hand, there is some evidence which suggests that extraverts do not perform as well as introverts in a learning setting (Entwistle and Wilson, 1978). However, there will obviously be other attributes which will influence a person's overall performance.

9.5 Summary

The analysis and interpretation reported here is an attempt to relate performance in the learning sessions in terms of the misunderstandings experienced, to a person's cognitive and learning styles and personality. A fundamental modelling difficulty arises because there are many facets to a person. Furthermore, each person is a combination of different styles, some may interact with and reduce the effects of others or alternatively accentuate an effect. It will only be through the testing of more people possessing different combinations of cognitive and learning styles and personality factors that a fuller picture can emerge and the isolation of components can be achieved. Until this is achieved it remains impossible to tell precisely the contribution of each factor to performance.

9.6 Day 3

The rationale for transcribing day 3 video tapes is that those commands that are "logical" to perform will be remembered better than those which do not follow the law of consistency. The extent of the transcripts are variable because some word processors are easier to interact with in terms of procedure.

All day 3 tapes were viewed and any difficulties noted. Where interesting dialogues emerged, these were transcribed. The classification of misunderstandings used for days 1 and 2 was employed. In addition, one extra category was included. However, this problem was only associated with one of the word processors. Control and shift key interference refers to errors occurring as a result of confusing these two keys.

A frequency count was not conducted as its usefulness would be minimal. It was thought that it would be sufficient to know which misunderstandings occurred for each word processor and how many of the subjects were affected under each category. From Table 9.3, one can see that the distribution and frequency of the subjects over the classification differ according to word processor. On day 3, WS still causes the greatest number of misunderstandings (11), compared to WP (8) and PW (9). However, the distribution of the subjects over the classification is less for PW than WP and WS.

9.7 End of session verbal protocols

At the end of each learning session, subjects spent a few minutes in an open interview with the experimenter. During these sessions, subjects were asked a number of questions. These included information on how the system was structured and how it operated, if there were any similarities between the word processors, any difficulties they found while interacting with the word processors and how much they liked the package. The protocols are not included in the thesis because they are long and consistency between the extent of the end of session protocols is low. In other

	C1	C2	C3
Basic editing problems	2	-	1
Incorrect command used	2	1	1
Locating a command	3	3	1
Did not read screen	1	2	1
How to carry out a command	2	3	1
Difficulty with menus	-	-	1
Don't know function of a command	1	2	1
Misunderstands function	1	2	1
Wrong cursor position	2	-	1
Typeover v's deletion	1	1	-
Inconsistent procedures	-	2	-
Incorrect spelling/case	1	-	-
Control & shift interference	1	-	-
Total	11	8	9

Table 9.3 Number of Subjects Displaying Misunderstandings on Day 3

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

words, some of the sessions induced a great deal of stress on the subjects. Thus I did not feel it was ethical to continue the study for longer than was necessary. A brief examination of the protocols does, however, suggest that subjects preferred WP over PW and WS, and none of the subjects liked WS. WS, in fact, was criticised most heavily. These included interference of the shift and control keys because they were too close together on the keyboard, that the system did not operate logically and they did not understand the terminology.

9.8 Discussion

The verbal protocols have permitted both qualitative and quantitative analyses of the video recordings. The qualitative analysis serves to highlight typical misunderstandings and how they occur, from which a misunderstanding classification was produced. On the other hand, a quantitative analysis permits comparisons between subjects' performance on the word processors, indicating ease of use. In addition, the quantitative analysis of the misunderstandings permitted an investigation of the misunderstandings which were predictable from the subject's personal style.

The findings suggest that it is possible to investigate personal style by observing the misunderstandings that occur in the learning sessions. The problem is that it is difficult to know precisely which attributes directly influence performance. The findings from the behavioural data (Chapter 8) have led to the refinement of the definitions and how personal style affects performance while these data suggest the types of misunderstanding and problems that users of particular personal styles may encounter. It may be possible to integrate the findings from the behavioural data and the quantitative analysis of the verbal protocols into a model of how cognitive and learning styles and personality affect performance on the basis of how much support a user requires within an interaction, and therefore by default, what sort of environment a user would require in order to work at an optimal level, and in terms of types of difficulties most likely to be encountered. An increase in subject number would permit different combinations of personal style to emerge and result in a finer-grained classification.

Finally, it is necessary to establish the usefulness of employing the frequency of misunderstandings as an indicator of which word processor is most compatible with a user. Misunderstandings are certainly a factor of performance, that is, one would expect that the more one understands about the word processor, the better one's performance. This seems perfectly logical but in reality comprehension appears to have a weak relationship with performance. That is, it is possible for a user to perform relatively well without understanding why. From the end of session verbal protocol of S3, af+er her first session with WS, it was apparent that she did not understand how the menu system operated but nevertheless, she performed better on this word processor than on the other two.

Excerpt from S3's end of session verbal protocol:

E: Can you tell me how the information is structured, how it works?
N: Well, there seems to be two screens. It keeps flicking into another one. But I've not seen that screen yet.
E: So you think there's two screens, menus like that?
N: But I don't know how to get from one to another and I haven't read it...

This is because the formula for performance is not simply understanding minus misunderstanding. The results of Study 4 have suggested that the relationship between performance and comprehension of a system's operations may be weak, and that it is still possible to perform well irrespective of understanding. Chapter 10 describes a study which explores further this relationship, in addition to examining any interactions with personal style.

If the supposition that misunderstanding is not the opposite of understanding is true then what other factors can be investigated which will permit differentiation of users? Another dimension which is still relatively unexplored is question-asking techniques. Over the last decade natural language systems have become popular, although they are now being superceded by multimodal interfaces. Natural language systems permit the user to communicate in their own language, although they still remain highly constrained. Pask's work on conversational style would suggest that language is related to learning style and hence to overall performance. There is also evidence that reflectives are better at asking questions which eliminate the alternative solutions in problem-solving tasks (Ault, 1973; Denney, 1973). Since we communicate via language, our problem solving techniques must be deeply ingrained in language. Study 5 examines the relationship between natural language and personal style.

9.9 Conclusions

It appears that it is possible to use cognitive and learning styles and personality factors to predict typical misunderstandings likely to occur in an interaction. However, the classification is still not sufficiently fine-grained and conclusive. Its use, however, might become more powerful when used in conjunction with the findings from Study 4. Further research needs to be conducted using a larger subject population and cross-testing of materials and style may prove useful.

Chapter 10 Personal Styles and Question-asking Techniques

10.1 Introduction

The results of Study 4 have suggested that there is a relationship between cognitive and learning styles and personality (personal style) and performance on different word processing packages. This was demonstrated by the behavioural analyses of the video tapes in particular, subjects' frequency of question asking, expert explanations and time on the manual were good indicators of personal style. It was also possible to predict from the subjects' personal style to which word processor each was best suited. This was corroborated by examining the data on number of moves taken, consulting the manual, consulting the expert and time on task. That is, subjects performed better on the word processor that was suited to their personal style.

In addition, the verbal protocols (in terms of misunderstandings) produced from these sessions also reflected differences in styles. This suggests that it is possible to differentiate between the way a person learns to use a computer according to one's personal style.

The results of Study 4 suggest it is possible to use many different types of information about an interaction and relate this to a user's personal attributes. Most of this information consisted of performance data and behavioural data of the subject's learning session while interacting with the expert. In particular the frequency of question asking was important. Study 5 attempts to relate the user's attributes to the subject's question-asking techniques. In one way this is similar to investigating natural language at the interface when confronted with a database query system. Intuition suggests that users of different styles will approach a problem differently and that this will be reflected in natural language. Furthermore, identification of a user's typical vocabulary and concept usage is one technique Rich (1983) has suggested for building user models.

10.2 Aims

The aims of Study 5 are:

i) to explore further the relationship between cognitive and learning styles and personality and performance, in particular question-asking techniques.

ii) to demonstrate differences in comprehension of the system's operations according to one's personal style. It is expected that the breadth of understanding of how the word processor operates will be reflected in the user's personal style. However, comprehension may not parallel performance (Study 4). An investigation of style and performance is still warranted and may suggest the characteristics which are most conducive to success.

10.3 Methodology

10.3.1 Subjects

13 postgraduates and research staff (7 males and 6 females)

voluntarily participated in this study. Subjects possessed varied amounts of computer experience.

10.3.2 Apparatus

Subjects learned to use the word processing package "Xchange" on the IBM PC-XT. Subjects received pre- and post-questionnaires which were designed by the author.

10.3.3 Procedure

The study was divided into 2 phases: phase 1 consisted of testing subjects individually on an identical battery of tests to Study 4 (EFT, MFFT, Divergent tests, a Styles of Learning Questionnaire and the EPI).

Phase 2 has 3 parts: pre-test, test and post-test. In the pre-test subjects are required to fill out visual analogue scales designed to elicit information about the subject's experience of computers (see Appendix 10.1). The test is a 25 minute session learning to use a word processor called "Xchange". Instructions to the subject and tasks to carry out can be found in Appendix 10.2. The text (Appendix 10.3) consisted of 2 paragraphs which had already been typed in. Subjects are told to make as many as possible of the specified alterations to the text and, second, to learn as much as possible about how the word processor operates. Subjects are told that they may ask questions and the experimenter responds with concise information, similar to the way an intelligent front end might respond. Over-cooperation is avoided and explanations are not given unless specifically asked for.

After the session, subjects filled in visual analogue scales for the degree of difficulty and similarity with other word processors with which the subject may be familiar. The post-test also involved questions to elicit the user's model of how the word processor operates.

10.4 Results

Each session lasted approximately 40 minutes and generated 3 hours and 25 minutes of video recordings. The types of data collected can be divided into:

- the battery of personal styles tests obtained by testing on an individual basis.
- questionnaire data consisting of:

i) visual analogues scales (VASs) designed to elicit the subject's level of computer experience, how much the subject likes computers, how difficult the subject found the word processor to operate and how similar the word processor was found to be to other word processors.

ii) comprehension questions designed to assess how much the subject understands about how the word processor operates Subjects were instructed to give as much detail as possible and if the subject did not know the answer they were permitted to guess.

- question-asking techniques obtained by analyses of the video tapes.

Questionnaire data: The mean value of the scores for all the visual analogue scales on computer experience was obtained (questions 1, 2b, 2c, 3a, 3b, 4).

Comprehension scores: Subjects received marks according to the complexity and accuracy of their answers. The maximum score possible was 40 and subjects scored between 3 and 32 points.

Number of tasks completed: The total number of tasks to be completed during the 25 minute session was 16. Subjects received a score according to the number of tasks that they were able to complete.

Video recordings: All video tapes were analysed by the author according to the types of questions asked and the frequency of consulting the help menu. Again, inter-rater reliability was not used because it was possible to accurately code the questions and behaviours. The types of questions recorded were:

- Confirmation questions; e.g., Is that right?
- What does X mean?; e.g., What does typeface mean?
- What now?; e.g., How do I proceed from here?
- How do I?; e.g., How do I load a document?
- Why?; e.g., Why did that happen?
- Do I/Can I/Is it?; e.g., Do I do it like this? Can I do it like this? Is it like this that I do it?
- Indirect questions; e.g., I'm not sure how to do that.

10.4.1 Personal Style and Comprehension

The battery of personal styles tests (EFT, MFFT, Styles of Learning, EPI and Divergent tests) were regressed against comprehension scores. Holistic orientation, the MFFT scores and inaccuracy on this test were found to be good predictors of comprehension scores $F(3,9)=10.91$, $p<0.0025$. Subjects who scored high on holistic orientation did better on the comprehension test than those subjects scoring low on holistic orientation. The holist learns by attempting to fit all the pieces of information that s/he accumulates about how the system operates into a general overview. They also have a tendency to use analogies to help them understand. The main feature of the redundant holist is that s/he personalises learning by constructing his/her own imaginary terms. Thus, redundant holists might benefit when learning time is exceptionally short. Impulsives also appear to have a better understanding than reflectives. The literature suggests the converse; that the reflective tends to have better comprehension because s/he reasons before making moves while the impulsive is prone to inaccuracy because s/he tends to "jump the gun". One reason for this counter-intuitive result is the time limit favours the impulsive who is concerned with working quickly. Therefore, the impulsive is able to complete more tasks and understand more. Reflectives may perform better in less constrained situations where no time limit is imposed. Impulsives tend to make more errors, thus impulsivity and inaccuracy are positively correlated. Hence, inaccuracy also emerged as a good predictor of comprehension scores.

In addition, it may be the case that impulsives ask questions more frequently and at an earlier stage than reflectives. Thus, there may be differences in how they utilise help facilities. Poor or primitive facilities presumably would not favour impulsives.

10.4.2 Experience and Comprehension

The mean score obtained for the set of VASs designed to elicit subjects' amount of computer experience and how much s/he likes computers was also found to be a good predictor of comprehension scores $F(1,11)=5.96$, $p<0.05$. The number of word processors with which the subject was already familiar was a significant predictor of comprehension $F(1,11)=11.65$, $p<0.01$ but the degree of similarity between this word processor and others with which the subject is familiar was not a significant predictor nor was perceived difficulty. As expected, the number of tasks the subject was able to complete affected their respective comprehension scores $F(1,11)=14.06$, $p<0.005$.

10.4.3 Question-asking Techniques and Comprehension

The different types of questions asked during the experimental session were not good predictors of comprehension scores. This may be due to the categories being too narrow. When the categories are reduced from 7 to 3 (confirmation questions, direct questions, indirect questions) with the addition of the category: frequency of help menu consultation and subjected to a multiple regression analysis with comprehension scores as the dependent variable, none of the question-asking techniques emerged as significant but frequency of help menu consultation appeared to be marginally correlated with comprehension scores ($t=2.092$, $p<0.1$). It appears that those subjects who consulted the menu more frequently also tended to have higher comprehension scores than those who did not. The total score for frequency of question-asking did not predict comprehension scores. Thus, high comprehension scorers did not differ in their question-asking techniques from low comprehension scorers.

10.4.4 Personal Style and Question-asking

Each part of the battery of personal styles tests was regressed against the full set of question-asking techniques, the reduced set plus frequency of help menu consultation and frequency of question-asking (summated score).

The EFT: Two of the question-asking techniques appeared to be marginally related to EFT scores. These were the "What now?" and "Why?" questions ($t=2.12$ and $t=2.54$, $p<0.1$) Field dependents appear to spend more time asking these types of questions in comparison to field independents. These results confirm those of Study 1 (map reading and EFT) which suggested that field dependents need more supportive information as to how to proceed. Therefore, they tended to consult the map more than field independents. "What now?" questions are vital for making the next move, hence this question is more frequently asked by field dependents than their counterparts. None of the questions in the reduced set or frequency of question-asking emerged as significant predictors.

Holistic orientation, holistic style, serialist style: the full and reduced set of questions and frequency of question-asking were not good predictors of holistic orientation, holistic style or serialist style.

Versatility: The frequency with which indirect questions are asked can significantly predict versatility $F(1,11)=11.65$, $p<0.01$. Versatile learners tend to ask more indirect questions. However, one might expect versatile learners to ask fewer of these questions since such a question-asking technique is not an effective way of asking. On the other hand, indirect questions may be more sophisticated, for example, children do not use this form frequently.

Pathologies: There are no significant relationships between pathologies and question-asking.

The MFFT: Impulsives might be expected to ask more questions than reflectives; this was suggested in Study 4. The results of this study suggest that no one particular type of question-asking technique can significantly predict MFFT scores nor are the reduced set or frequency of question-asking good predictors of reflectivity-impulsivity.

Divergent thinking: Question-asking techniques are not good predictors of divergent thinking.

Extraversion: Question-asking techniques are not good predictors of extroversion.

Stable: Question-asking techniques are not good predictors of stability in a person but the frequency of question-asking is marginally significant as a predictor of stability $F(1,11)=4.366$, $p=0.06$. As would be expected less stable users tend to ask more questions. One type of question that might possibly be expected to predict stability is confirmation questions but this relationship did not emerge.

The findings from Study 4 suggested that impulsives ask more questions than reflectives. However, instability could also have caused a higher frequency of question-asking since the two subjects who were impulsives were also unstable. This study suggests that it is not impulsivity that leads to increased question-asking but instability.

10.4.5 Experience and Question-asking

Experience: "How do I?" questions emerged as significant predictors of the level of experience $F(1,11)=11.27$, $p<0.01$. The more experienced a user is, the less frequently this question occurs. As the level of experience increases, the user will take a more active role in human computer interaction. Therefore, questions such as "how do I?" will become less frequent because the more experienced user has a better model of how the word processor operates. One would also expect that as the level of experience increases, the frequency of question-asking would decrease. This was confirmed $F(1,11)=4.83$, $p<0.05$.

The number of word processors with which the subject is familiar can be predicted from the frequency with which s/he consults the help menu $F(1,11)=6.19$, $p<0.05$. This indicates that those subjects who are familiar with other word processors consulted the help menu on a more frequent basis. This may be due to the fact that these subjects knew that a help facility existed and were much better equipped to use it. This argument is corroborated by the fact that the number of times the help menu was consulted is a good predictor of level of experience $F(1,11)=12.05$, $p<0.01$).

The number of word processors with which the subject is familiar can be predicted by the frequency of question-asking $F(1,11)=5.71$, $p<0.05$. Those subjects who were familiar with other word processors asked fewer questions overall.

10.4.6 Number of Tasks Completed and Question-asking

Question-asking techniques do not appear to predict the number of tasks each subject was able to complete in the 25 minute session. One might have expected that strategic question-asking might lead to more useful information and thus influence the ability to complete the tasks.

10.4.7 Number of Tasks Completed and Personal Style

Holistic orientation, holistic style and instability were found to be good predictors of the number of tasks that the subject was able to complete $F(3,9)=13.58$, $p<0.002$. Comprehension learners tend to be able to complete more of the tasks, in particular redundant holists. On the other hand, the findings from the question-asking techniques do not suggest that redundant holists and holists ask more effective questions or have a higher frequency of question-asking. It may be the case that redundant holists and holists fortuitously happened to be the more experienced computer users. However, inspection of the correlations demonstrates that there is a weak relationship between holistic orientation and experience and holistic style and experience. This suggests that there may be a component of the holistic learning that makes these students more successful at learning to use computers which is not reflected in the help they might receive. They may be more favoured in short learning sessions because they are able to understand the workings of the computer in a more global fashion. They also use past experience and analogies to help them understand how all the components fit into a framework. The redundant holist is able to invent his/her own personal descriptions and fill in the parts that are missing in the framework. Operation learners (serialist style) work in a step-by-step manner and only integrate the accumulate information when they are forced to do so. Thus, they are not able to see immediately how the word processor operates at a global level.

Instability also predicted the number of tasks completed. Subjects who were unstable completed more tasks than stable subjects. Unstable subjects also asked more questions, thus they were able to proceed more quickly. One possible explanation is that the unstable subject users were more influenced by the experimental setting and the time limit imposed, despite the fact that the instructions did not specifically stress that they were to complete all the tasks set.

10.5 Discussion

Using multiple regression analyses, it has been possible to investigate which factors are necessary for effective learning in the domain of human computer interaction. The findings of this study suggest that the most efficient learner in terms of the number of tasks completed and understanding of how the system operates, is the redundant holist. "... of the redundant holist is that imaginary descriptive terms are used... what seems to happen is that the redundant holist personalises learning" (Entwistle and Ramsden, 1983). The holist tends to look ahead using many analogies and attempting always to integrate the new information into a broader view. On the other hand, serialists progress in a more cautious step by step manner and do not integrate the incoming information until forced to do so. Are serialist or operation learners less successful than holist or comprehension learners in learning to use word processors and if so why? Designers of systems have intentionally exploited the use of analogies to improve ease of learning. In this particular study, time is limited so serialists are able to gather pieces of information but do not get the opportunity to integrate the information into a larger framework. Additionally, serialists were in general slower at performing the tasks than holists so the number of tasks they were able to complete in the time limit was fewer, and this in turn also affected their comprehension scores. One might expect the versatile learner who possesses both types of traits to be most advantaged. This might well be the case in a natural learning setting where learning is open-ended and timeless. However, in a longer study (cf. study 4) comprehension learners were also found to be the more successful learners of word processors.

It was proposed that the types of questions the subject asks might indicate when a more effective learning strategy is being used. The set of question-asking techniques did not appear to be a major contributor to successful learning. For example, in general, personal style could not be predicted from the types of questions asked. There were a few exceptions such as the relationship between versatility and indirect questions. The frequency of question-asking tended to be a little more useful: unstable and less experienced subjects were more indulgent.

The VASs designed to elicit experience and the number of previous word processors used were good measures of ability to learn a new word processor. However, some personal style traits also indicated a subject's ease of learning. For example, the redundant holist and holist appear to be the most successful learners. Other aspects which emerged were the relationship between impulsivity and comprehension and instability and number or tasks completed. It is possible that these results merely reflect the design of the study.

What has been learned about user's understanding of how word processors operate from this study? Study 4 suggested that a weak relationship exists between performance and comprehension. Performance in this case refers to the number of moves, time on task and other behavioural data. Since performance data are not identical in Study 5, there are problems of comparability. Performance in this study refers to the number of tasks completed and the number of questions asked. In these terms, comprehension learners out-performed operation learners in terms of the number of tasks completed and also demonstrated a superior understanding of the system's operations. This may be due to the interaction of comprehension and the number of tasks completed.

10.6 Conclusions

The results of the study suggest that learning and some types of cognitive styles and personality inventories may be an alternative approach to investigating and building self-adaptive interfaces. The results also suggest that particular styles and personality are related to certain types of behaviour. The utilisation of question-asking techniques to investigate the types of problems a user might encounter has not been particularly fruitful.

Chapter 11 Concluding Remarks

This thesis has attempted to fill some gaps in the area of human computer interaction, especially in relation to adaptive interfaces. It has also taken a rather unusual methodological approach to understanding what is going on in an interaction by using both qualitative and quantitative approaches to exploring the data including applied and naturalistic settings.

Two major themes run through the thesis. They are:

i. the use of colour as a means of support to the user

ii. the construction of an adaptive interface through exploration of the relationship between personal style and performance on computer tasks.

Study 1 began by exploring the relationship between the EFT and performance on a computer task and varying the amount of support available to the subject. Subjects' performance differed in terms of performance measures such as the number of moves taken to complete each task component, frequency of 'help' consultation and time on task. From EFT scores it was possible to differentiate between subjects' performance. The performance of field independent subjects tended to be superior to that of field dependent subjects, and, this margin of superiority was maintained on day 2. An alternative explanation offered for this difference in performance was that some subjects had poorer memory than others. A study (Study 2) was constructed which specifically examined memory for colour using similar parameters to Study 1. The two tasks also differ: the task in Study 1 required the subject to manipulate information spatially and to remember particular pieces of information, while the task in Study 2 required the subject to remember associations between forms of coding and location by creating mnemonics for remembering. The two tasks may tap different cognitive components. The similarity between them is that in Study 1, subjects in the support condition (colour) can augment their performance by decreasing memory load by associating the colour with their position on the map. In Study 2, the task of association is more overt. The results of Study 1 suggest that subjects allocated to the support condition did not attempt to make the associations and no interaction of type of support (colour or no colour) and cognitive style emerged. One would expect that irrespective of cognitive style, subjects who used the colour as a support would be able to augment their performance. Some literature suggests that subjects must be made aware of the usefulness of the colour coding (e.g., Eriksen, 1952). The instructions of this study asked subjects to use the colour to help them know where they were on the map but did not explicitly point out its usefulness. This may have contributed to the lack of significant findings but more probable is the fact that colour needs to be used carefully and that colour can easily be overused and therefore hinder performance.

In Study 2 when subjects were forced to make associations between the type of cue (position, colour or shape), both colour and shape were found to be effective cues in relation to remembering the position. The expected effect of style and sex on associative learning emerged but there may have been sampling problems.

A second aim of Study 2 was to investigate the usefulness of different "cues" to the user. The results of this study suggested that colour used in redundant non-qualitative manner is no better a feedback cue than shape coding. Both forms of coding were more helpful than no coding at all. However, processing load was higher for the no coding condition. This does not deny the value of the study because this is precisely the problem with which users of menu-type systems are confronted. That is, users are often confronted with hierarchies of menus with no indication as to where they are. They are not provided with an overview of the hierarchies on line but with part of the architecture labelled in some form.

Study 3 continued to explore the use of colour coding at the interface and reports an investigation of the use of non-redundant qualitative colour, while Studies 4 and 5 reported further investigations of adaptive interfaces by exploring the relationship between personal styles and performance on word processors.

Study 3 investigated the qualitative colour coding of facial expressions. The idea of capturing non verbal gestures by image processor (see Sheehy and Forrest, 1986b) and using facial expressions and postures to give feedback to the user (see Sheehy et al, 1988) have been explored elsewhere. Most of the research on colour has focussed on non-qualitative colour coding of meaningless symbols. This study set out to explore the effects of inherent knowledge on performance. The findings suggest that there are too many intrasubject differences despite the fact that the colours supposedly convey qualitative, intrinsic meaning. In addition, people have individual preferences for particular colours.

Studies 1, 2 and 3 employed slightly artificial approaches to studying human computer interaction. Study 1 described a computer task and similarities are drawn between the map reading task and learning to use a word processor. This parallel may appear artificial and it may be claimed that the relationship between the tasks is weak. The idea of the building metaphor arose from work conducted as part of an ESPRIT programme of research and is an extension of the desktop metaphor (see Rogers, Leiser and Carr, 1988). The building metaphor entails complexity and may induce problems with "getting lost" in the building. Although Rogers et al found that subjects understood the building metaphor better than the desktop or system metaphor, this may not be the case when a comparison of performance measures are conducted in a naturalistic setting. Colour coding parts of the building could reduce the complexity of the metaphor as well as cognitive loads. In exploring the building metaphor and its capabilities, Studies 1 and 2 emerged.

These studies have attempted to investigate complex problems using simple paradigms, while Studies 4 and 5 take naturalistic approaches to investigating the interface by observing how users interact with word processors.

Study 4 is a "case study" design. A small n was chosen in order that the study be in-depth. Additionally, the vast quantity of data that would emerge from a larger subject sample might restrict the extent of the data analyses. In Study 4 all video tapes were analysed for the three subjects on several dimensions, both quantitatively and qualitatively, and the findings were related to the subject's personal style. Case study designs are already widely used in clinical settings and intrasubject differences are considered important because of the reasons described in Studies 2 and 3.

Perhaps more importantly, researchers should not be "bullied" into using large subject samples because of tradition and in hope of gaining more meaningful results. Large subject samples necessitate the measurement of ever decreasing effects, in addition, to a waste of resources (Lunt and Livingstone, 1988). In some subject areas it might be more natural to employ single case subject designs because intrasubject differences are important or the case is unique for each subject. Study 4 employs a small subject number precisely because human computer interaction entails one human user interacting with one computer.

The cognitive and learning styles and personality of each person is different, if only in subtle ways. There will be some people who will have similar cognitive and learning styles and they can be expected to perform similarly. The three subjects studied showed similarities and differences and it was possible to isolate some of the factors which tend to produce particular types of behaviour such as frequency of asking questions. The study also illustrated the problems of measurement and discrimination that are likely to confront dynamic adaptive interfaces.

Small subject sizes mean that traditional statistical techniques cannot be used because many of the assumptions are violated e.g., independence of scores. Case studies tend to rely on visual inspection of the data or time series analysis. Naturally it is only possible to use the latter type of statistics when the data permit. There are alternative statistics, such as ANOVAs, which have been developed for use with small n. However, their usefulness have been criticised (e.g., Hartmann, 1974; Peck, 1985). This meant that only tentative conclusions could be drawn from visual inspection of the data. A multiple correlation matrix was computed but the small subject size necessitated extremely high correlational values (0.997) in order to reach significance levels. We must assume that both Type 1 and 2 errors may occur if patterns in the data cannot be established.

The remainder of the data were subjected to visual inspection and predictions were made of how subjects should perform according to their personal style. These predictions were largely supported. Another method of coding similarities and dissimilarities of the subjects' performances which was used was based on types of misunderstandings that they encountered. In some sense this type of coding is similar to Cantor, Brown and Groat's (1985) Multiple Sorting Procedure, except that the manner in which the categories are sorted is not subjective but objective.

In general, Study 4 has indicated how one might expect people of different styles to behave and it has also helped to refine the definitions of these different styles. There is however some room for improvement such as employing a larger subject population of people differing widely in their personal style. Instead of employing case study designs, a larger subject size would permit the use of traditional statistical analyses. Statistical techniques could be used to compare the performances of the subjects in relation to their personal styles rather than to group them and therefore lose the importance of intra-subject differences. The problem with quantitative analyses is precisely the manner in which the intrasubject differences are "flattened out".

An increase in subject size would also be useful in constructing a table of dominance to indicate how the different styles interact with each other to increase or reduce a particular behaviour.

In addition, day 3 could be made a continuation of days 1 and 2 so that day 3 could be an extension of their learning rather than a test of previously learned material. The subject's ability to continue learning new material may rest on what they have already learned, that is, their "model" of how the word processor operates. This information would prove more useful as a test of comprehension and performance.

Study 4 produced a wealth of data yet it was not a particularly lengthy investigation. Other studies (e.g., Mack et al, 1983) have suggested that novice computer users need approximately 8 hours to learn just the basic editing skills. There is a need for longer studies but the quantity of data such studies will produce probably acts as a deterrent to researchers and designers alike.

Lastly, the study was purposely less constrained than others have been in the past. This has been useful but again raises problems of analysis. It might have been useful to have employed a structured interview at the end of each session to give better comparability. Again, one must consider the merits of open interviews yielding more qualitative data and the merits of structured interviews which tend to produce quantitative data. One compromise might be the use of repertory grids (Kelly, 1955) or multiple sorting procedures (MSP) (Cantor et al, 1985).

Study 5 takes up some of the issues addressed in Study 4: it employs a larger number of subjects but the learning phase is shorter (25 minutes). Again, it might have been more profitable if the learning phase was followed by a "continuation" learning phase. This would have enabled a more accurate picture of the user's model of how the system operates to have been built up and reduce the importance of the part played by short term memory factors. Here too, the employment of multiple sorting procedure or repertory grids may have been useful in depicting the user's understanding of the system's operations and would also eliminate the possibility that the learning was short term and superficial.

The main focus of Study 5 was to explore further the relationship between personal style and performance through investigation of question-asking techniques. It has been suggested that reflectives are better at asking the necessary questions which quickly eliminate incorrect solutions (Ault, 1973; Denney, 1973). Study 5 demonstrated that it was not impulsivity but instability that correlated positively with frequency of question-asking.

The study also explored style and comprehension. Study 4 suggested that a weak relationship exists between performance and comprehension and therefore it was necessary to examine what type of person was likely to understand more about the system's operations and what type of person was likely to perform well. However, Study 4 viewed performance and comprehension as identical and it was only through the open interview at the end of the sessions that it was apparent that performance did not always reflect comprehension. Study 5 used a separate comprehension test after the learning phase and it was found that comprehension learners (in particular, redundant holists) understood most about the word processor's operations but also completed more of the tasks. It is not clear whether comprehension learners had better understanding of the word processor because their personal style was more conducive to success or because they were able to complete more tasks in the 25 minutes. Despite this, high comprehension scorers did not differ in their question-asking techniques from low comprehension scorers nor did they ask more questions in general. There is some evidence that serialists might ask questions that hinder performance.

Comprehension learners were found to consult the help menu more frequently as were more experienced users. Thus it is not clear whether comprehension learners are better equipped to learn word processing than other types of learners or whether this result should be attributable to the fact that they were on the whole more experienced with computers. Visual inspection of the data suggests that comprehension learners were not always the more experienced subjects.

The results from this study are inconclusive and better types of performance measures are required. In addition, the study might have been more useful if all the subjects had completed all the tasks and if time had been used as a performance measure. The comprehension scores might then have been more comparable between different types of users.

The studies, taken as a whole, have shown the following:

i. Some types of users need more support than other types and users would benefit from different environments in which to work. Users can be categorised by field dependence.

ii. Colour is one way of providing feedback to the user about his/her location in a menu-driven system. However, when the user is learning, demands on cognitive processing is high. The ability to use colour as a support might be hindered during the early stages of learning especially if the colour is too specific. Other forms of achromatic coding are equally effective.

iii. It may not be possible to use colour in a qualitative non-redundant manner because of intrasubject differences.

iv. Users of different personal styles interact differently with word processing systems and the "expert". Users perform according to their personal style.

v. Users of different styles encounter different types of problems which can be investigated by observing the forms and origins of the misunderstandings they experience.

vi. Some types of question-asking techniques may be related to the personal style of the user but not to overall comprehension of the system's operations.

vii. There is not a simple relationship between comprehension and performance because of the nature of action dependencies in complex tasks.

Future Research

Techniques for the investigation of individual differences have involved repertory grid (Kelly, 1955) and multiple sorting procedure (Cantor et al, 1985). These techniques are specifically useful for investigating conceptual systems. In particular, the MSP is sensitive to individual differences without the procedural disadvantages of the rep. grid. For example, the grid limits the number of elements that can be handled and the overall time to complete the grid may be high (Cantor, Brown and Richardson, 1976). Using the MSP, it might be possible to study comprehension in a more in-depth and meaningful way (cf. Folley and Williges, 1982). In addition, it might be possible to investigate how novices become expert users over time. More research needs to be conducted on a long term basis. In particular, if cognitive and learning styles are relatively static then it might be interesting to see at what point in the learning phase that field dependents catch up with field independents in terms of performance. Alternatively, some types of users may be more predisposed to becoming "expert" users rather than experienced users.

The MSP also permit a classification of users depending on how users sort the elements. For example, serial field independents may sort on a different basis to global field dependents. This may be a good way of understanding each type of user's conceptual system, in addition to exploring the bases for misunderstandings.

It is also possible to explore similarities and dissimilarities in the workings of different software packages to show how particular users have difficulty understanding certain concepts. This information could then be used by system's designers to make systems more usable. The MSP would also permit comparisons to be made of the designer's model and the user's model. Thus the MSP can be used as an evaluation tool. The larger the discrepancy between the designer's and user's sorting of elements, the more incompatible the product (interface) is to the user and hence give rise to difficulties using the product.

Hudgens and Billingsley's (1978) article "Sex: the missing variable" did not result in promoting sex as an important variable to be studies or accounted for in research (cf. Fowler and Murray, 1988). The literature on cognitive styles such as the field dependence-independence and reflectivity-impulsivity dimensions have demonstrated that there are sex differences in both the way males and females perceive and analyse objects and the way in which they respond to many stimuli. If these differences exist, then sex is a factor that needs to be studied especially in relation to the interface design of automated office systems and other word processors. Presently, design is male dominated, and design is based on how males perceive women or perhaps in terms of how males perceive "users" in general. Interfaces need to be adaptive purely because users are not of one type but can be categorised on a number of dimensions, sex being just one of them.

Bibliography

Adcock, C.J. and Webberley, M. (1971) Primary Mental Abilities. **Journal of General Psychology, 84, 229-243.**

Allen, J.F. and Perrault, C.R. (1980) Analyzing Intention in Utterances. **Artificial Intelligence, 15, 143-178.**

Allsopp, J.F. and Eysenck, H.J. (1974) Personality as a Determinant of Paired-associates Learning. **Perception and Motor Skills, 39, 315-324.**

Alty, J.L. (1985) Use of Path Algebras in an Interactive Dialogue System. **Interact '84**, B. Shakel, (Ed.). North -Holland: Elsevier Science Publishers.

Ault, R.L. (1973) Problem Solving Strategies of Reflective, Impulsive, Fast-accurate and Slow-inaccurate Children. **Child Development, 44, 259-266.**

Baer, D.M. (1977) "Perhaps It Would Be Better Not to Know Everything". **Journal of Applied Behavior Analysis, 10,** 167-172.

Bakan, D. (1954) A Generalization of Sidman's Results on Groups and Individual Functions and a Criterion. **Psychological Bulletin, 51, 63-64.**

Bariff, M.L. and Lusk, E.J. (1977) Cognitive and Personality Tests for the Design of Management Information Systems. **Management Science, 23, 820-829.**

Benbasat, I. and Dexter, A.S. (1979) Value and Events Approaches to Accounting: An Experimental Evaluation. **The Accounting Review, 54,734-749.**

Benyon, D. (1985) MONITOR. A Self-adaptive User Interface. In **Interact '84**, B. Shakel (Ed.). North-Holland: Elsevier Science Publishers.

Benyon, D., Innocent, P., Shergill, J. and Murray, D. (1986) Experiments in Adaptive Interfaces. **Colloquium on Adaptive Man-Machine Interfaces, IEE, Computing and Control Division.**

Berger, E. (1977) Field Dependence and Short Term Memory. Doctoral Dissertation, New York University. **Dissertation Abstracts International 38, no.1870B (University of Microfilms no. 77-12, 266.**

Bijl, A. and Szalapaj, P. (1985) Saying What You Want with Words and Pictures. **Interact '84**, B. Shakel, (Ed.). North-Holland: Elsevier Science Pubilshers.

Block, J., Block, J.H. and Harrington, D.M. (1974) Some Misgivings about the Matching Familiar Figures Test as a Measure of Reflection-Impulsivity. **Developmental Psychology, 10, 611-632.**

Bond A. and Lander M. (1974) The Use of Analogue Scales in Rating Subjective Feelings. **British Journal of Medical Psychology, 47, 211-218.**

Bone, R.N. (1971) Interference, Extraversion and Paired-associate Learning. **British Journal of Social and Clinical Psychology, 10, 284-285.**

Breakwell, G.M., Fife-Shaw, C., Lee, T. and Spencer, J. (1986) Attitudes to New Technology in Relation to Social Beliefs and Group Memberships: A Preliminary Investigation. **Current Psychological Research and Reviews, 5, 34-47.**

Cameron, R. and Myers, J.L. (1966) Some Personality Correlates of Risk-taking. **Journal of General Psychology, 74, 51-60.**

Campbell, S.B. and Douglas, V.I. (1972) Cognitive Styles and Responses to the Threat of Frustration. **Canadian Journal of Behavioural Science, 4, 30-42.**

Cantor, D., Brown, J. and Richardson, H. (1976) Constructs without Tears: Is there Life beyond the Grid. Paper presented to British Psychological Society Annual Conference, Exeter.

Cantor, D., Brown, J. and Groat, L. (1985) A Multiple Sorting Procedure for Studying Conceptual Systems. In **Research Interview: Uses and Approaches**, M. Brenner, J. Brown and D. Cantor (Eds.). London: Academic Press.

Carroll, J.M. and Mack, R.L. (1984) Learning to Use a Word Processor: By Doing, by Thinking and by Knowing. In **Human Factors in Computer Systems**, J.C. Thomas and M.L. Schneider (Eds.). Ablex Publishing Company.

Carroll, J.M. and Thomas, J.C. (1982) Metaphor and the Cognitive Representation of Computing Systems. **IEEE Transactions on Systems, Man and Cybernetics, 12, 107 -116.**

Carter, R.C. (1979) Visual Search with Color. Doctoral dissertation, Pennsylvania State University.

Carter, R.C. (1982) Visual Search with Color. **Journal of Experimental Psychology: Human Perception and Performance, 8, 127-136.**

Chernoff, H. (1973) The Use of Faces to Rrepresent Points in k-dimensional Space Graphically. **Journal of the American Statistical Association, 68, 361-368.**

Christ, R.E. (1975) Review and Analysis of Color Coding Research for Visual Displays. **Human Factors, 17, 542 -570.**

Christ, R.E. (1984) Research for Evaluating Visual Display Codes: An Emphasis on Colour Coding. In **Information Design**, R. Easterby and H. Zwaga (Eds.). Chichester:John Wiley.

Christ, R.E. and Corso, G.M. (1983) The Effects of Extended Practice on the Evaluation of Visual Display Codes. **Human Factors, 25, 71-84.**

Christner, C.A. and Ray, H.W. (1961) An Evaluation of the Effect of Selected Combinations of Target and Background Coding on Map-Reading Performance: Experiment V. **Human Factors, 3, 131-146.**

Cohen, P.R. and Perrault, C.R. (1979) Elements of a Plan -based Theory of Speech Acts. **Cognitive Science, 3, 177 -212.**

Cooper, M. and Hockley, A. (1986) The Evaluation of an Adaptive Interface. **Colloquium on Adaptive Man-Machine Interfaces, IEE, Computing and Control Division.**

Cooper, P.A. (1986) The Adaptive Intelligent Dialogues Project- Phase 1 Summary. **Colloquium on Adaptive Man -Machine Interfaces, IEE, Computing and Control Division.**

Davidoff, J.B. (1986) The Role of Colour in Object Recognition, Categorisation and Naming. In **Contemporary Ergonomics**, D.J. Oborne (Ed.).

Davidoff, J.B. (1987) The Role of Colour in Visual Displays. In **International Reviews of Ergonomics, Current Trends in Human Factors Research and Practice, volume I**, D.J. Oborne (Ed.). London: Taylor Francis.

Denney, D.R. (1973) Reflection and Impulsivity as Determinants of Conceptual Strategy. **Child Development, 44, 614-623.**

DeProspero, A. and Cohen, S. (1979) Inconsistent Visual Analyses and Presentation of Graphic Data. In **Single Subject Research Strategies for Evaluating Change**, T.R. Kratochwill (Ed.). New York: Academic Press.

Edmonds, E. (1974) A Process for the Development of S.W. for Non-technical Users as an Adaptive System. **General Systems, 19, 215-217.**

Edmonds, E. (1981) Adaptive Man-Computer Interfaces. In **Computing Skills and the User Interface**, M.J. Coombs and J.L. Alty (Eds.). London: Academic Press.

Edmonds, E. (1986) Towards a Taxonomy of User Interface Adaptation. **Colloquium on Adaptive Man-Machine Interfaces, IEE Computing and Control Division.**

Edmonds, E. and Guest, S.P. (1978) SYNICS - A FORTRAN Subroutine Package for Translation. **Report Number 6, Man-Computer Interface Research Group, Leicester Polytechnic.**

Entwistle, N.J. (1978) Knowledge Structures and Styles of Learning: A Summary of Pask's Recent Research. **British Journal of Educational Psychology, 48, 255-265.**

Entwistle, N.J. and Ramsden, P. (1983) Understanding Student Learning. London: Croom Helm.

Entwistle, N.J. and Wilson, J.D. (1970) Personality, Study Methods and Academic Performance. **University Quarterly, 21, 147-166.**

Entwistle, N.J. and Wilson, J.D. (1977) Degrees of Excellence: The Academic Achievement Game. London: Hodder and Stoughton.

Ericsson, K.A. and Simon, H.A. (1980) Verbal Reports as Data. **Psychological Review, 87, 215-251.**

Ericsson, K.A. and Simon, H.A. (1985) Protocol Analysis, Verbal Reports as Data. MIT: Cambridge, Mass.

Eriksen, C.W., (1952) Location of Objects in a Visual Display as a Function of the Number of Dimensions on which the Object Differ. **Journal of Experimental Psychology, 44, 55-60.**

Eysenck, H.J. (1967) The Biological Basis of Personality. Springfield: C.C. Thomas.

Eysenck, H.J. (1970) The Structures of Human Personality. London: Methuen.

Eysenck, H.J. (1971) Editor's Introduction. In **Readings in Extraversion-introversion, Vol. 3.** London: Staples.

Eysenck, M.J. (1977) Human Memory: Theory, Research and Individual Differences. Oxford: Pergamon Press.

Farmer, E.W. and Taylor, R.M. (1980) Visual Search through Color Displays: Effects of Target-Background Simularity and Background Uniformity. **Perception and Psychophysics, 27, 267-272.**

Flemming, M.L. (1968) Message Design: The Temporal Dimension of Message Structure. Cited in D.R. Goodenough (1976) The Role of Individual Differences in Field Dependence as a Factor in Learning and Memory. **Psychological Bulletin, 83, 675-694.**

Folley, L.J. and Williges, R.C. (1982) User Models of Text Editing Command Languages. **Human Factors in Computing Systems, 326-340.**

Fowler, C., Macauley, L. and Fowler, J. (1986) The Relationship between Cognitive Style and Dialogue Style: An Exploratory Study. In **Proceedings of the HCI '85 Conference.**

Fowler, C. and Murray, D. (1988) Gender and Cognitive Style Differences at the Human Computer Interface. In **Interact '87**, H.J. Bullinger and B. Shackel (Eds.). North Holland: Elsevier Science Publishers.

Frank, B.M. (1983) Flexibility of Information Processing and the Memory of Field Independent and Field Dependent Learners. **Journal of Research in Personality, 17, 89-96.**

Gaines, B.R. and Shaw, L.G. (1983) Dialog Engineering. In **Designing for Human Computer Interaction.** Academic Press.

Geiselman, R.E. and Samet, M.G. (1982) Notetaking and Comprehension for Computer-Displayed Messages: Personalised versus Fixed Formats. **Human Factors in Computing Systems, 45-50.**

Gentile, J.R., Roden, A.H. and Klein, R.D. (1972) An Analysis-of-Variance Model for the Intrasubject Replication Design. **Journal of Applied Behavior Analysis, 5, 193-198.**

Gilfoil, D.M. (1982) Warming upto Computers: A Study of Cognitive and Affective Interaction over Time. **Human Factors in Computing Systems, 245-250.**

Goodenough, D.R. (1976) The Role of Individual Differences in Field Dependence as a Factor in Leaning and Memory. **Psychological Bulletin, 83, 675-694.**

Greenberg, S. and Witten, I.H. (1985) Adaptive Personalised Interfaces - A Question of Viability. **Behaviour and Information Technology, 4, 31-45.**

Hanne, K.H., Hoepelman, S.Ph. and Fahnrich, K.P. (1986) Combined Graphics/Natural Language Interfaces to Knowledge Based Systems. Paper presented at Esprit Technical Week, Brussels, September 1986.

Harris, L.R. (1977) User Oriented Database Query with the ROBOT N.L. Query System. **International Journal of Man -Machine Studies, 9, 697-713.**

Hartmann, D.P. (1974) Forcing Square Pegs into Round Holes: Some Comments on "An Analysis-of-Variance Model for the Intrasubject Replication Design". **Journal of Applied Behavior Analysis, 7, 635-638.**

Hartmann, D.P., Gottman, J.M., Jones, R.R., Gardner, W., Kazdin, A.E. and Vaught, R.S. (1980) Interrupted Time -series Analysis and Its Application to Bahavioural Data. **Journal of Applied Behavior Analysis, 13, 543-559.**

Hasher, L. and Zacks, R.T. (1979) Automatic and Effortful Processes in Memory. **Journal of Experimental Psychology: General,108, 356-388.**

Hayes, S.C. (1981) Single Case Experimental Design and Empirical Clinical Practice. **Journal of Consulting and Clinical Psychology, 49, 193-211.**

Hays, W.L. (1963) Statistics for Psychologists. New York: Holt, Rinehart and Winston.

Herman, H.J.M. (1970) A Questionnaire Measure of Achievement Motivation. **Journal of Applied Psychology, 54, 353-363.**

Hill, A.J. (1987) Investigation of some Short Term Influences on Hunger, Satiety and Food Consumption in Man. Unpublished Doctoral Thesis. University of Leeds.

Hiltz, S.R. and Turoff, M. (1982) Human Diversity and the Choice of Interface: A Design Challenge. **ACM SIGSOC Bulletin, 13, 125-130.**

Hodges, W.F. and Spielberger, C.D. (1969) Digit Span: An Indicant of Trait or State Anxiety? **Journal of Consulting and Clinical Psychology, 33, 430-434.**

Hoe, R.V., Poupeye, K., Vandierendonck, A. and de Soete, G. (1987) Some Effects of Menu Characteristics and User Personality on Performance with Menu-driven interfaces. **International Conference on Communication and Data Communication.**

Hudgens, G.A. and Billingsley, P.A. (1978) Sex: the Missing Variable in Human Factors Research. **Human Factors, 20, 245-250.**

Innocent, P.R. (1982) Towards Self-Adaptive Interface Systems. **International Journal of Man-Machine Studies, 16, 287-299.**

Jensen, A. (1964) Individual Differences in Learning: Interference Factor. **Cooperative Research Project No. 1867, Office of Education, U.S. Dept. of Health Education and Welfare.**

Jones, R.R., Vaught, R.S. and Weinrott, M. (1977) Time -series Analysis in Operant Research. **Journal of Applied Behavior Analysis, 10, 151-166.**

Jones, R.R., Weinrott, M.R. and Vaught, R.S. (1978) Effects of Serial Dependency on the Agreement between Visual and Statistical Inference. **Journal of Applied Behavior Analysis, 11, 277-283.**

Kagan, J. (1966) Reflection-Impulsivity: The Generality and Dynamics of Conceptual Tempo. **Journal of Abnormal Psychology, 71, 71-24.**

Kagan, J., Rosman, B.L., Day. D., Albert, J. and Phillipa, W. (1964) Information Processing in the Child: Significance of Analytic and Reflective Attitudes. **Psychological Monographs, 78, Whole No.578.**

Kamin, L.J. and Fedorchak, O. (1957) The Taylor Scale, Hunger, and Verbal Learning. **Canadian Journal of Psychology, 11, 212-218.**

Karat, J. (1986) Transfer between Word Processing Systems. **In Proceedings of the International Scientif Conference: Work with Display Units, Part II, 745-748.** Stockholme, May 12-17.

Katkin, E.S. (1965) Relationship between Manifest Anxiety and Two Indices of Autonomic Stress. **Journal of Personality and Social Psychology, 2, 324-333.**

Kelly, G. (1955) The Psychology of Personal Constructs. New York: Norton.

Kidd, A. (1982) Man-Machine Dialogue Design. **British Telecom Reports, 1, Study 1.**

Klemmer, E.T. and Lockhead, G.R. (1962) Productivity and Errors in Two Keying Tasks: A Field Study. **Journal of Applied Psychology, 46, 401-408.**

Kratochwill, T.R., Alden, K., Demuth, D., Dawson, D., Panicucci, C., Arntson, P., McMurray, N., Hempstead, J. and Levin, J. (1974) A Further Consideration in the Application of an Analysis-of-Variance Model for the Intrasubject Replication Design. **Journal of Applied Behavior Analysis, 7, 629-633.**

Lovaas, O.I. (1960) The Relationship of Induced Muscular Tension, Tension Level and Manifest Anxiety in Learning. **Journal of Experimental Psychology, 59, 145-152.**

Luder, C.B. and Barber, P.J. (1984) Redundant Color Coding in Airbourne CRT Displays. **Human Factors, 26, 19-32.**

Lunt, P. and Livingstone, S. (1988) Psychology and Statistics: Testing the Opposite of the Idea You First Thought Of. Paper presented at BPS Social Section, Annual Conference, Canterbury, September 1988.

Luria, S.M., Neri, D.F. and Jacobsen, A.R. (1986) The Effects of Set Size on Colour Matching Using CRT Displays. **Human Factors, 28, 49-61.**

Lusk, E.J. and Kersnick, M. (1979) The Effect of Cognitive Style and Report Format on Task Performance: The MIS Design Consequences. **Management Science, 25, 787-798.**

Macauley, L. and Norman, M. (1985) Designing Interfaces for Different Types of Users - Experimental Criteria. In **Interact '84**, B. Shakel (Ed.). North-Holland: Elsevier Science Publishers.

Mack, R.L., Lewis, C.H. and Carroll, J.M. (1983) Learning to Use Word Processors: Problems and Prospects. **ACM Transaction on Office Information Systems, 1, 254-271.**

Martindale, C. and Greenough, J. (1973) The Differential Effect of Increased Arousal on Creative and Intellectual Performance. **Journal of Genetic Psychology, 123,** 329-335.

Maskery, H.S. (1985) Adaptive Interfaces for Naive Users - An Experimental Study. **Interact '84**, B. Shakel (Ed.). North-Holland: Elsevier Science Publishers.

McCain, L.J. and McCleary, R. (1979) The Statistical Analysis of the Simple Interrupted Time Series Quasi -experiement. In **Quasi-experimentation: Design and Analysis Issues for Field Settings**, T.D. Cook and D.T. Campbell (Eds.). Chicago: Rand mcNally.

McClean, P.D. (1968) Paired-associate Learning as a Function of Recall Interval, Personality and Arousal. Unpublished PhD Thesis, London.

Messer, S.B. (1976) Reflection-Impulsivity: A Review. **Psychological Bulletin, 83, 1026-1052.**

Miller, L.A. and Thomas, J.C. jr. (1977) Behavioural Issues in the Use of Interactive Systems. **International Journal of Man-machine Studies, 9, 509-536.**

Minsky, M. (1975) A Framework for Representing Knowledge. In **The Psychology of Computer Vision**, P.H. Winston (Ed.). New York: McGraw-Hill.

Mosher, F.A. and Hornsby, J.R. (1966) On Asking Questions. In **Studies in Cognitive Growth**, J.S. Bruner, R.R. Olver and P.M. Greenfield (Eds.). New York: Wiley.

Mulgrave, N.W. (1965) An Investigation of the Cognitive Factor Structure of 2 Types of Concept-attainment Tasks and 2 Tests of Cognitive Style. Doctoral Dissertation, University of Pittsburgh. **Dissertation Abstracts 27, 1283B (University of Microfilms, no. 66-8158).**

Murch, G.M. (1985) Colour Graphics - Blessing or Ballyhoo? **Computer Graphics Forum, 4, 127-135.**

Narborough-Hall, C.S. (1985) Recommendations for Applying Colour Coding to Air Traffic Control Displays. **Displays, Technology and Applications, 6, 131-137.**

Neimark, E.D. (1975) Longitudinal Development of Formal Thought. **Genetic Psychology Monographs, 91, 171-225.**

Nievergelt, J. (1982) Errors in Design and how to avoid Them. In **Document Preparation Systems: A Collection of Survey Articles**, J. Nievergelt, G. Coray, J.D. Nicoud and A.C. Shaw.

Nisbett, R.E. and Wilson, T.D. (1977) Telling More Than We Can Know: Verbal Reports on Mental Processes. **Psychological Review, 84, 231-259.**

Nuessle, W. (1972) Reflectivity as an Influence on Focusing Behavior of Children. **Journal of Experimental Child Psychology, 14, 265-276.**

Ogden, W.C. and Boyle, J.M. (1982) Evaluating Human -computer Dialog Styles: Command Versus Form/Fill-in for Report Modification. In **Proceedings of the Human Factors Society - 26th Annual Meeting, 542-545.**

Osborne, J.W. (1972) Short- and Long-term Memory as a Function of Individual Differences in Arousal. **Perception and Motor Skills, 34, 587-593.**

Ostergaard, A.L and Davidoff, J.B. (1975) Some Color Effects on Naming and Recognition of Objects. **Journal of Experimental Psychology: Learning, Memory and Cognition, 11, 579-587.**

Parsonson, B.S. and Baer, D.M. (1978) The Analysis and Presentation of Graphic Data. In **Single Subject Research Strategies for Evaluating Change**, T.R. Kratochwill (Ed.). New York: Academic Press.

Pask, G. and Scott, B.C.E. (1972) Learning Strategies and Individual Competence. **International Journal of Man -machine Studies, 4, 217-253.**

Pask, G. et al (1977) Third Progress Report on SSRC Research Programme, HR 2708, Systems Research Limited, Richmond, Surrey.

Peck, D.F. (1985) Small N Experimental Designs in Clinical Research. In **New Developments in Clinical Psychology**, F.N. Watts (Ed.). Chichester: John Wiley & Sons.

Potosnak, K.M. (1986) Interface Design for Different Types of Computer Users. In **Proceedings of the International Scientific Conference: Work with Display Units, Part II**, 586-589. Stockholme, May 12-17.

Ramond, C.K. (1953) Anxiety and Task as Determiners of Verbal Performance. **Journal of Experimental Psychology, 46**, 120-124.

Ramsey, H.R. and Grimes, J.D. (1983) Human Factors in Interactive Computer Dialog. **Annual Review of Information, Science and Technology (ARIST), 18**, 29-59.

Rappaport, H. and Katkin, E.S. (1972) Relationships among Manifest Anxiety, Response to Stress and the Perception of Autonomic Activity. **Journal of Consulting and Clinical Psychology, 38**, 219-224.

Reali, N. and Hall, V. (1970) Effect of Success and Failure on the Reflective and Impulsive Child. **Developmental Psychology, 3**, 392-402.

Reising, J.M. and Emerson, T.J. (1985) Colour in Quantitative and Qualitative Display Formats. **Journal of Institution of Electronic and Radio Engineers, 55**, 397 -400.

Rich, E. (1979) Using Modeling via Stereotypes. **Cognitive Science, 3**, 329-354.

Rich, E. (1983) Users are Individuals: Individualising User Models. **International Journal of Man-Machine Studies, 18**, 199-214.

Robinson, J.A. and Bennink, C.D. (1978) Field Articulation and Working Memory. **Journal of Research in Personality, 12**, 439-449.

Rogers, Y., Leiser, R. and Carr, D. (1988) Evaluating Metaphors for Use at the User-system Interface. **Proceedings of Eurinfo '88**.

Sackman, H. (1970) Experimental Analysis of Man-computer Problem-solving. **Human Factors, 12**, 187-201.

Schank, R.C. and Abelson, R.P. (1977) Scripts, Plans, Goals and Understanding. N.J.: LEA.

Shapiro, M.B. (1966) The Single Case in Clinical -psychological Research. **The Journal of General Psychology, 74, 3-23.**

Sheehy, N.P. and Forrest, M-A. (1986b) Esprit Project 527 Communication Failure in Dialogue, Deliverable 5.

Sheehy, N.P. and Forrest, M-A. (1986a) Esprit Project 527 Communication Failure in Dialogue, Deliverable 6.

Sheehy, N.P., Forrest, M-A. and Chapman, A.J. (1988) Conversing with Data-bases in Automated Office Environments. In J.C. Jesuino and G.M. Stephenson (Eds.) **Social and Environmental psychology in the European Context.** Amsterdam: Nijhof.

Sidman, M. (1952) A Note on Functional Relations Obtained from Group Data. **Psychological Bulletin, 49, 263-269.**

Siegal, A.W., Kirasic, K.C. and Kilburg, R.R. (1973) Recognition Memory in Reflective and Impulsive Preschool Children. **Child Development, 44, 651-656.**

Siegelman, E. (1969) Reflective and Impulsive Observing Behavior. **Child Development, 40, 1213-1222.**

Spence, K.W. (1958) A Theory of Emotionally Based Drive (D) and Its Relation to Performance in Simple Learning Situations. **American Psychologist, 13, 131-141.**

Spiker, A., Rogers, S.P. and Cicinelli, J. (1986) Selecting Colour Codes for a Computer-Generated Topographic Map Based on Perception Experiments and Functional Requirements. **Ergonomics, 29, 1313-1328.**

Standish, R.R. and Champion, R.A. (1960) Task Difficulty and Drive in Verbal Learning. **Journal of Experimental Psychology, 59, 361-365.**

Stewart, T.F.M. (1974) Ergonomic Aspects of Man-Computer Problem-Solving. **Applied Ergonomics, 5, 209-212.**

Taylor, J.A. (1958) A Personality Scale of Manifest Anxiety. **Journal of Abnormal and Social Psychology, 57, 55-60.**

Van der Veer, G.C., Tauber, M.J., Waern, Y. and van Muylwijk, B. (1985) On the Interaction between System and User Characteristics. **Behaviour and Information Technology, 4, 289-308.**

Van Muylwijk, B., Van der Veer, G.C. and Waern, Y. (1983) On the Implications of User Variability in Open Systems. An Over-view of What We Know and the Lot We Have to Find Out. **Behaviour and Information Technology, 3, 313-326.**

Walther G.H. and O'Neil, H.F. Jr. (1974) On-line User -computer Interface - the Effect of Interface Flexibility, Terminal Type and Experience on Performance. **AFIPS Conference Proceedings, 43, 379-384.**

Williges, R.C. and North, R.A. (1973) Prediction and Cross-validation of Video Cartographic Symbol Location Performance. **Human Factors, 15, 321-336.**

Witkin, H.A., Dyke, R.B., Faterson, H.F., Goodenough, D.R. and Karp, S.A. (1962) Psychological Differentiation. New York: John Wiley.

Witkin, H.A., Ottman, P.K., Raskin, E. and Karp, S.A. (1971) Manual for EFT, Childrens' EFT and GEFT. Palo Alto Clalifornia: Consulting Psychologists Press.

Witkin, H.A., Moore, C.A., Goodenough, D.R. and Cox, P. (1977) Field Dependent and Field Independent Cognitive Styles and their Educational Implications. **Review of Educational Research, 47, 1-64.**

Witkin, H.A. and Goodenough, D.R. (1981) Cognitive Styles: Essence and Origin. New York: International University Press.

Wolins, L. (1982) Research Mistakes in the Social and Behavioral Sciences. Iowa: Iowa Sate University Press.

Zwaga, H.J.G. and Duinhower, F. (1984) The Influence of a Fixed Background in a VDU Display on the Efficiency of Colour and Shape Coding. In **Colour Coded vs Monochrome Electronic Displays, Proceedings of a NATO Workshop,** C.P Gibson (Ed.). London: HMSO.

A P P E N D I C E S

Appendix 3.1 Microtext Program to Display Building

```
................................................
*1
$TRACE SUMMARY
$BOX 4,15,25,70
Your task is to find the shortest route between
the Psychology department and the Psychology
laboratories. You are now in Psychology
Do you want to go through the door?

Your options are:
yes (press key 1)
no  (press key 2)
?
1=2
2=1
 =1
................................................
*2
This is the Chemistry department

Your options are:
Take door 1 (press key 1)
Take door 2 (press key 2)
Take door 3 (press key 3)
?
1=1
2=3
3=4
................................................
*3
You are in West 1 corridor

Your options are :
Turn left (press key 1)
Turn right (2)
Take door in front of you (3)
Turn around (4)
?
1=5
2=53
3=52
4=2
................................................
*4
You are in North corridor

Your options are:
Turn left (1)
Turn right (2)
Take door in front of you (3)
Turn around (4)
?
1=11
2=12
3=10
4=2
................................................
*5
You are at Major Intercept

Your options are:
Turn left (1)
Turn right (2)
Continue (3)
Turn around (4)
?
1=6
2=8
3=7
4=32
................................................
*6
You are in North corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=2
2=10
3=11
4=12
................................................
*7
You are in East 1 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=13
2=14
3=15
................................................
*8
You are in South 1 corridor
```

Your options are:
Continue (1)
Turn around (2)
?
1=17
2=16
....
#9
You are in North corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=10
2=2
3=12
4=11
....
#10
This is the Maths department

Your options are:
Take door 1 (1)
Take door 2 (2)
?
1=18
2=19
....
#11
You are at staircase 1.

You must turn around.

Please press spacebar
!
=9
....
#12
You are at Major Intercept

Your options are:
Turn left (1)
Turn right (2)
Continue (3)
Turn around (4)
?
1=7
2=32
3=8
4=6
....
#13
This is the French department

There is no other exit.

Please press spacebar to continue.
!
=23
....
#14
You are in East 1 corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=10
2=24
3=26
4=27
....
#15
You are at Major Intercept

Your options are:
Turn left (1)
Turn right (2)
Continue (3)
Turn around (4)
?
1=8
2=6
3=32
4=7
....
#16
You are at Major Intercept

Your options are:
Turn left (1)
Turn right (2)
Continue (3)
Turn around (4)
?

```
1=32
2=7
3=6
4=8
.................................................
*17
You are at West 2-East 2 intercept

Your options are:
Turn left (1)
Turn right (2)
Turn around (3)
?
1=44
2=36
3=25
.................................................
*18
You are in North corridor

Your options are:
Take door in front of you (1)
Turn left (2)
Turn right (3)
Turn around (4)
?
1=2
2=12
3=11
4=10
.................................................
*19
You are in East 1 corridor

Your options are:
Turn left (1)
Turn right (2)
Take door in front of you (3)
Turn around (4)
?
1=26
2=27
3=24
4=10
.................................................
*23
You are in East 1 corridor

Your options are:
Turn left (1)
Turn right (2)
?
1=15
2=14
.................................................
*24
This is the German department

There is no other exit.

Please press spacebar to continue.
!
=47
.................................................
*25
You are in South 1 corridor

Your options are:
Continue (1)
Turn around (2)
?
1=16
2=17
.................................................
*26
You are in East 1 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=50
2=49
3=51
.................................................
*27
You are in East 1 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=13
2=15
3=14
.................................................
*32
You are in West 1 corridor
```

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=52
2=2
3=53
4=5
....
#36
You are in West 2 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=45
2=56
3=57
....
#44
You are at South 2-East 2 intercept

Your options are:
Turn right (1)
Continue (2)
Turn around (3)
?
1=60
2=90
3=62
....
#45
This is the Biology department

Your options are:
Take door 1 (1)
Take door 2 (2)
?
1=52
2=54
....
#47
You are in East 1 corridor

Your options are:
Turn left (1)
Turn right (2)
Take door in front of you (3)
?
1=27
2=26
3=10
....
#49
You have reached the end of the corridor

You must turn around.

Please press spacebar to continue.
!
=63
....
#50
This is the womens' W.C.

There is no other exit.

Please press spacebar to continue.
!
=64
....
#51
You are in East 1 corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=24
2=10
3=27
4=26
....
#52
This is the Physics department

Your options are:
Take door 1 (1)
Take door 2 (2)
?
1=66
2=45
....

#53
You are in West 1 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=67
2=68
3=69
....
#54
You are in West 2 corridor

Your options are:
Turn left (1)
Turn right (2)
Turn around (3)
?
1=57
2=56
3=45
....
#56
You are in West 2 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=70
2=71
3=72
....
#57
You are in West 2-East 2 intercept

Your options are:
Turn left (1)
Continue (2)
Turn around (3)
?
1=25
2=44
3=36
....
#60
You are in South 2 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=73
2=61
3=95
....
#61
You are at South 2-East 3 intercept

Your options are:
Turn left (1)
Continue (2)
Turn around (3)
Turn around (3)
?
1=76
2=77
3=78
....
#62
You are at West 2-East 2 intercept

Your options are:
Turn right (1)
Continue (2)
Turn around (3)
?
1=25
2=36
3=44
....
#63
You are in East 1 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=50
2=51
3=49
....
#64
You are in East 1 corridor

```
Your options are:
Turn left (1)
Turn right (2)
?
1=51
2=49
................................................
#66
You are in West 1 corridor

Your options are:
Turn left (1)
Turn right (2)
Take door in front of you (3)
Turn around (4)
?
1=53
2=5
3=2
4=52
................................................
#67
This is the English department

There is no other exit.

Please press spacebar to continue.
!
=83
................................................
#68
You are at the end of the corridor.

You must turn around.

Please press spacebar to continue.
!
=100
................................................
#69
You are in West 1 corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
?
1=2
2=52
3=5
................................................
#70
This is the Art department

There is no other exit.

Please press spacebar to continue.
!
=85
................................................
#71
You have reached the end of the corridor.

You must turn around.

Please press spacebar to continue.
!
=88
................................................
#72
You are in West 2 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=45
2=57
3=56
................................................
#73
This is the cafeteria.

There is no other exit.

Please press spacebar to continue.
!
=87
................................................
#76
You are in East 3 corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=97
```

```
2=96
3=98
4=99
.................................................
#77
You are in South 2 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=106
2=107
3=111
.................................................
#78
You are in South 2 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=73
2=95
3=61
.................................................
#80
This is the mens' W.C.

There is no other exit.

Please press spacebar to continue.
!
=103
.................................................
#81
You are at the end of the corridor

You must turn around.

Please press spacebar to continue.
!
=93
.................................................
#83
You are in West 1 corridor

Your options are:
Turn left (1)
Turn right (2)
?
1=68
2=69
.................................................
#85
You are in West 2 corridor

Your options are:
Turn left (1)
Turn right (2)
?
1=86
2=71
.................................................
#86
You are in West 2 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=45
2=57
3=56
.................................................
#87
You are in South 2 corridor

Your options are:
Turn left (1)
Turn right (2)
?
1=95
2=61
.................................................
#88
You are in West 2 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=70
2=86
3=71
.................................................
```

```
*89
You are in South 2-East 2 intercept

Your options are:
Turn left (1)
Continue (2)
Turn around (3)
?
1=60
2=62
3=90
................................................
*90
You are in East 2 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=80
2=81
3=89
................................................
*93
You are in East 2 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=80
2=89
3=81
................................................
*95
You are at South 2-East 2 intercept

Your options are:
Turn left (1)
Turn right (2)
Turn around (3)
?
1=62
2=90
3=60
................................................
*96
You have reached the Psychology laboratories
Well done!

Please press spacebar
!
=127
................................................
*97
This is the Computing room

There is no other exit.

Please press spacebar to continue.
!
=113
................................................
*98
You are in East 3 corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=114
2=115
3=116
4=122
................................................
*99
You are at South 2-East 3 intercept

Your options are:
Turn left (1)
Turn right (2)
Turn around (3)
?
1=77
2=78
3=76
................................................
*100
You are in West 1 corridor

Your options are:
Take door on right (1)
Continue (2)
Turn around (3)
?
1=67
```

2=69
3=68
....
*103
You are in East 2 corridor

Your options are:
Turn left (1)
Turn right (2)
?
1=81
2=89
....
*106
This is the Library

There is no other exit.

Please press spacebar to continue.
!
=109
....
*107
You are at staircase 2

You must turn around.

Please press spacebar to continue.
!
=110
....
*109
You are in South 2 corridor

Your options are:
Turn left (1)
Turn right (2)
?
1=111
2=107
....
*110
You are in South 2 corridor

Your options are:
Take door on left (1)
Continue (2)
Turn around (3)
?
1=106

2=111
3=107
....
*111
You are at South 2-East 3 intercept

Your options are:
Turn right (1)
Continue (2)
Turn around (3)
?
1=76
2=78
3=77
....
*113
You are in East 3 corridor

Your options are:
Turn left (1)
Turn right (2)
Take door in front of you (3)
?
1=98
2=99
3=96
....
*114
This is the Language Laboratory

There is no other exit.

Please press spacebar to continue.
!
=121
....
*115
This is the History department

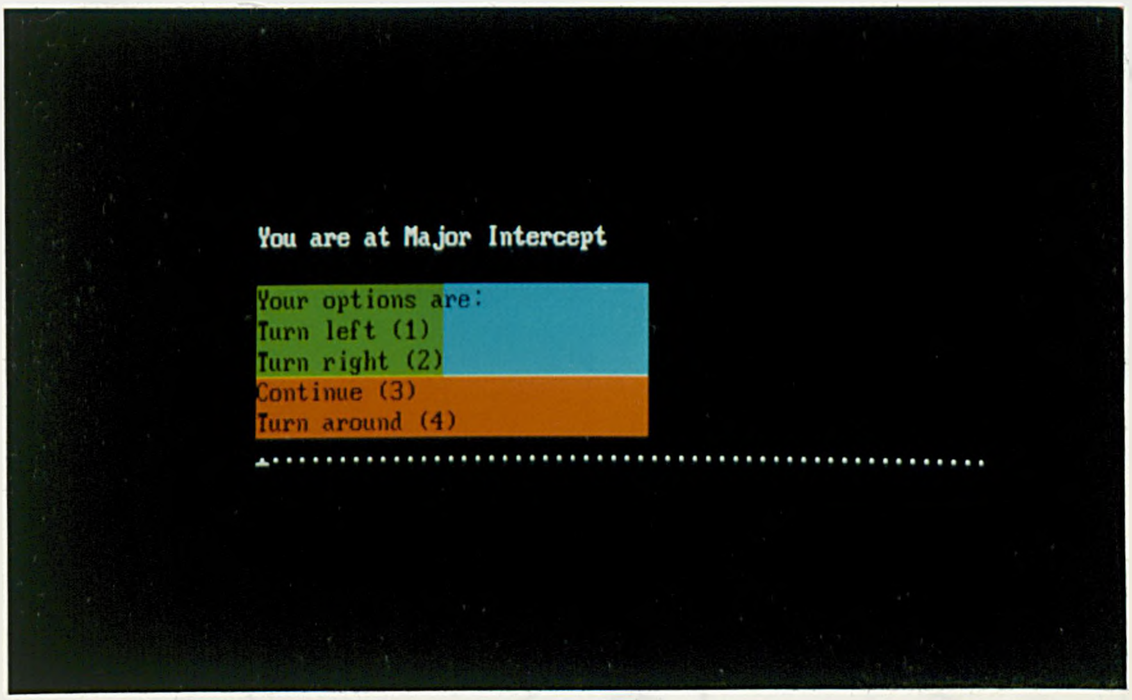
There is no other exit.

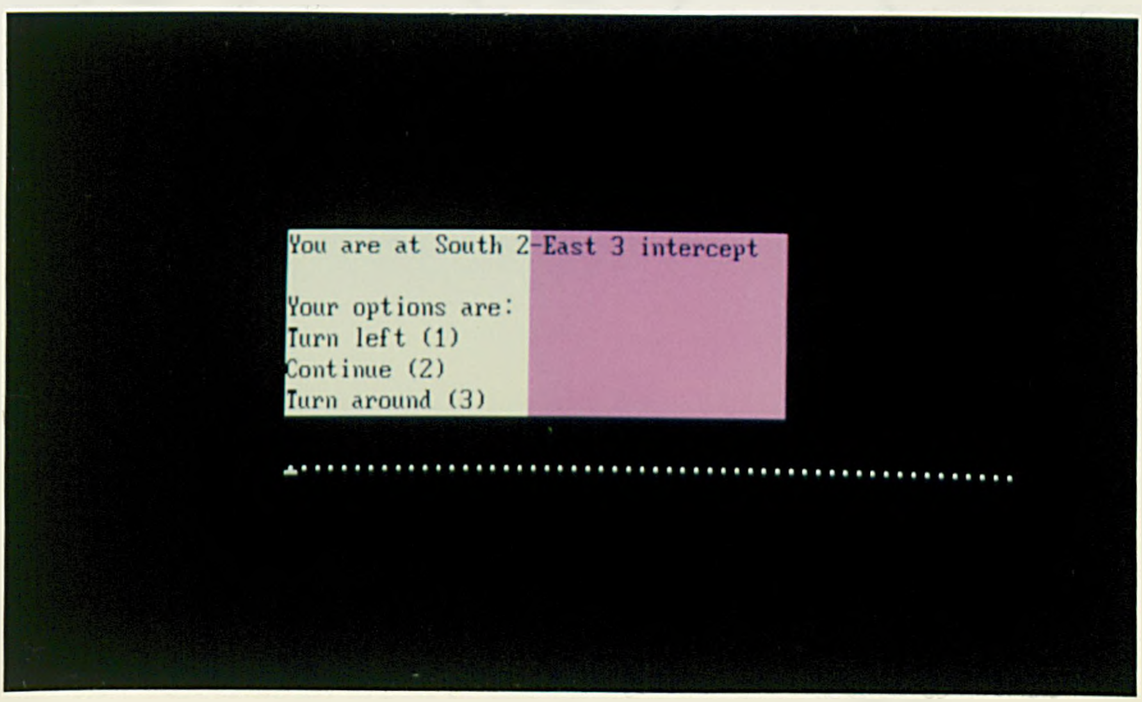
Please press spacebar to continue.
!
=117
....
*116
You are at the end of the corridor.

You must turn around.

```
Please press spacebar to continue.
!
=123
................................................
*117
You are in East 3 corridor

Your options are:
Turn left (1)
Turn right (2)
Take door in front of you (3)
?
1=122
2=116
3=114
................................................
*121
You are in East 3 corridor

Your options are:
Turn left (1)
Turn right (2)
Take door in front of you (3)
?
1=116
2=122
3=115
................................................
*122
You are in East 3 corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=96
2=97
3=99
4=98
................................................
*123
You are in East 3 corridor

Your options are:
Take door on left (1)
Take door on right (2)
Continue (3)
Turn around (4)
?
1=115
2=114
3=122
4=116
................................................
*127
$SUMMARY SAVE, MAF-C1.M $C
$END
................................................
```

Appendix 3.2 Photographs of Directions in Building





Appendix 5.1 Materials for Pilot Study: Faces and Adjectives

Page 1

ADJECTIVES

Fiendish
Embarrassed
Angry
Grumpy
Bored
Jealous
Apprehensive
Furious
Afraid
Sad
Happy
Fed up

Listed below are the 12 adjectives from page 1. Beside each adjective write down a colour which you think best represents that adjective. You may use shades of colour such as dark or light.

Fiendish ____________________

Embarrassed ____________________

Angry ____________________

Grumpy ____________________

Bored ____________________

Jealous ____________________

Apprehensive ____________________

Furious ____________________

Afraid ____________________

Sad ____________________

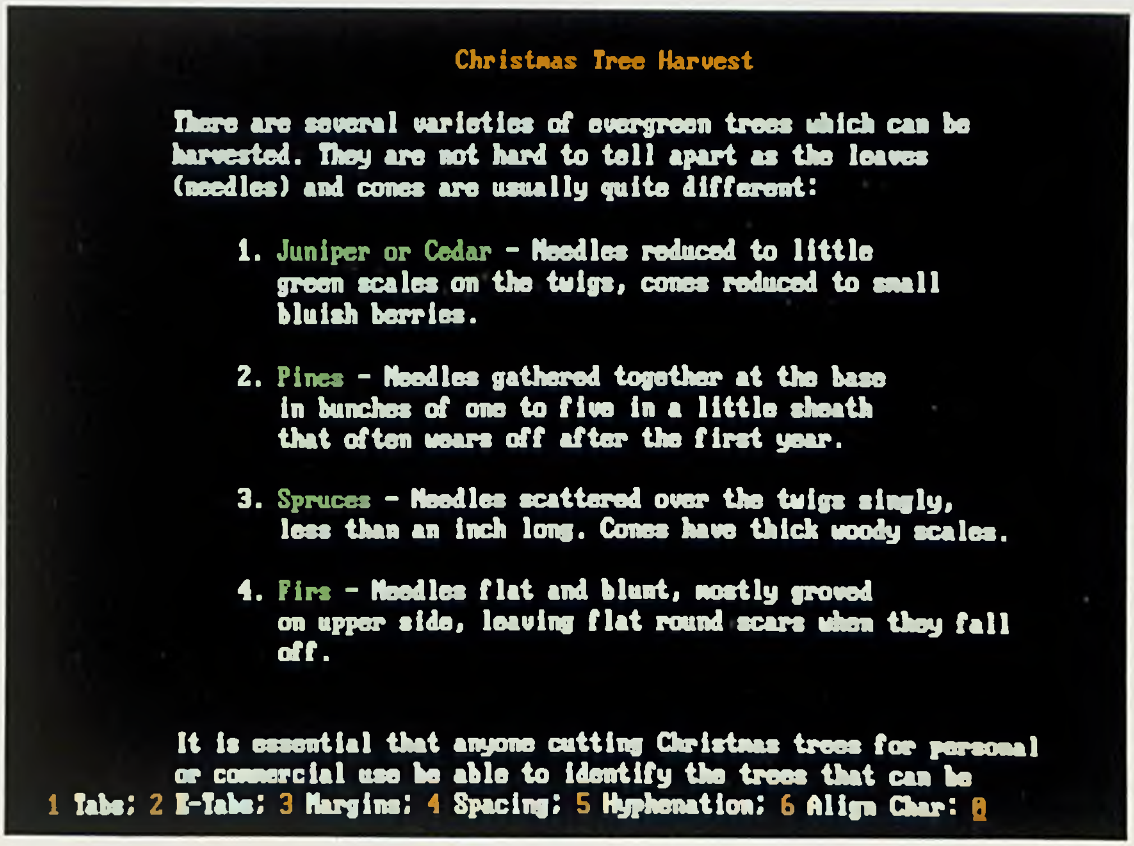
Happy ____________________

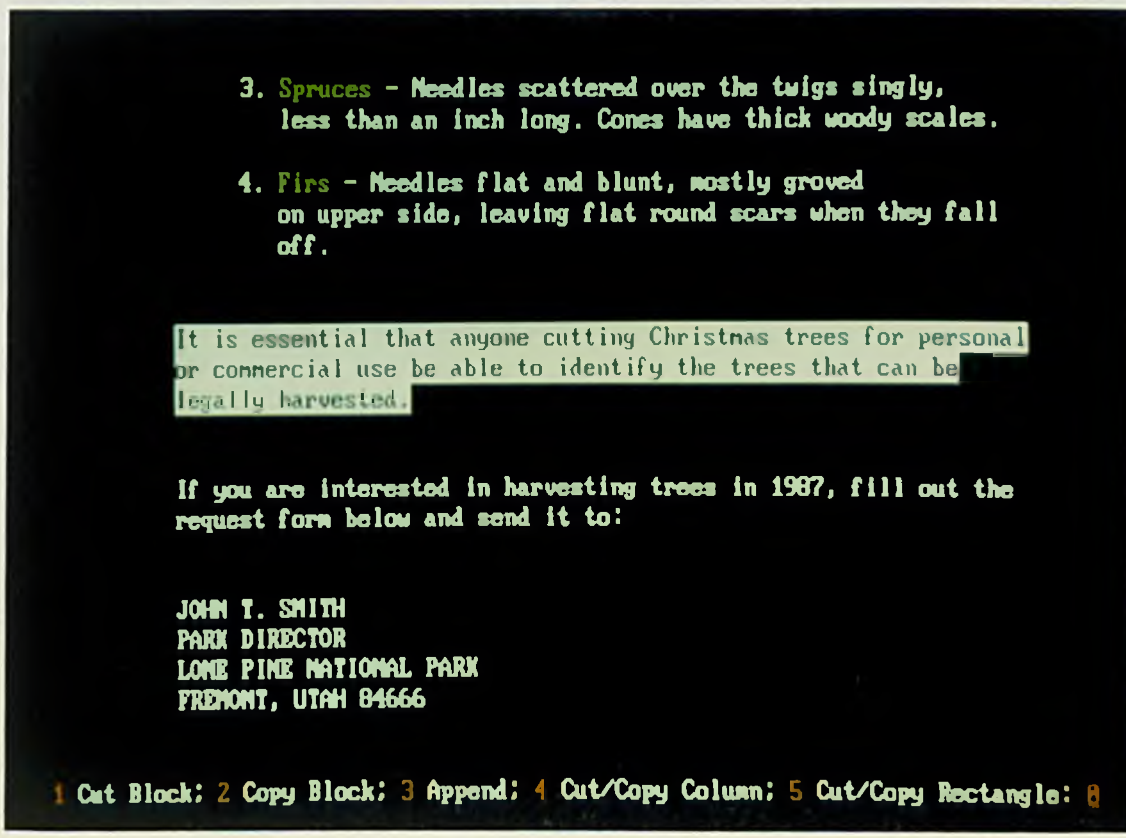
Fed up ____________________

Appendix 5.2 Program to Display Adjectives on Screen

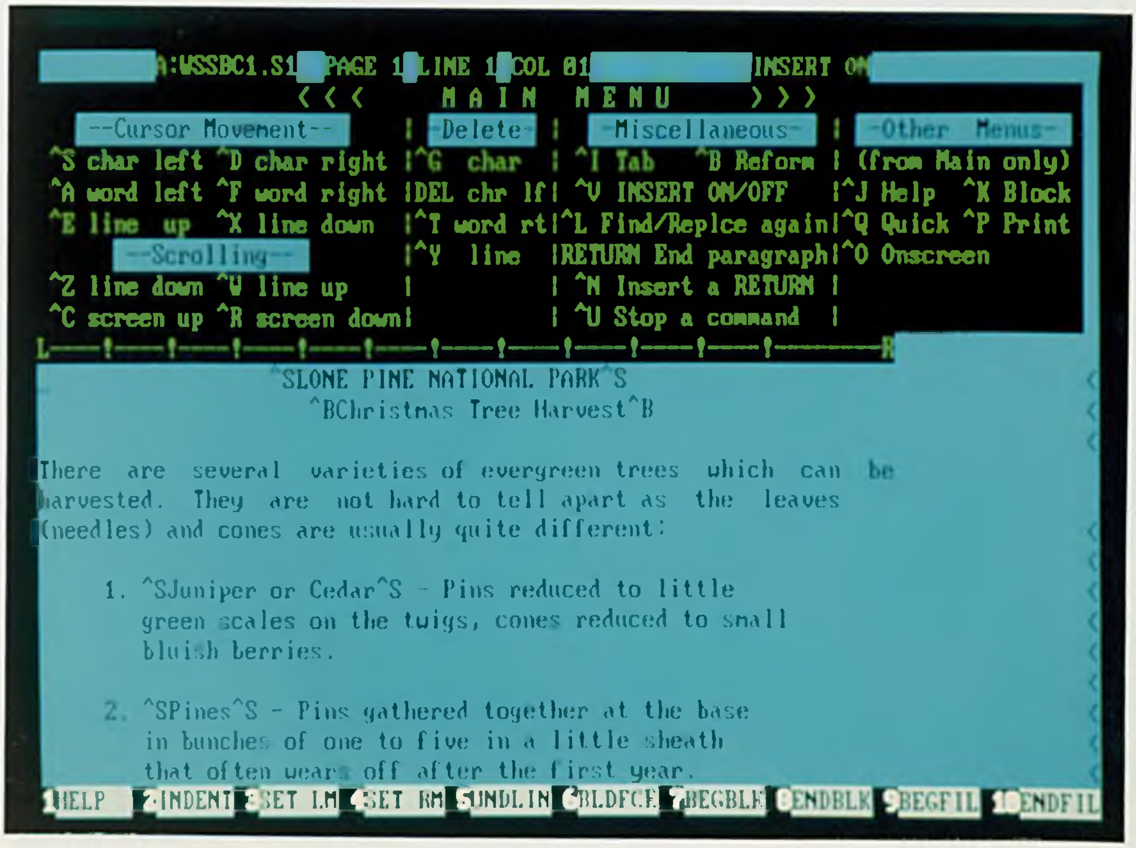
```
10 REM **THIS PROG DISPLAYS FACES**
20 DIM GESTURE$(40)
30 FOR X=1 TO 40
40 READ GESTURE$(X)
50 NEXT X
60 KEY OFF
70 FOR I=1 TO 40
80 CLS
90 LOCATE 12,38
100 PRINT GESTURE$(I)
110 OPEN "LPT1:" AS #1
120 TIME$="00:00:00"
130 A$=INPUT$(1)
140 PRINT #1, TIME$; TIMER
150 CLOSE
160 GOSUB 190
170 NEXT
180 END
190 CLS
200 FOR J=1 TO 2000
210 NEXT J
220 RETURN
230 DATA BORED
240 DATA HAPPY
250 DATA FURIOUS
260 DATA ANGRY
270 DATA HAPPY
280 DATA SAD
290 DATA EMBARRASSED
300 DATA SAD
310 DATA FED UP
320 DATA BORED
330 DATA JEALOUS
340 DATA EMBARRASSED
350 DATA FED UP
360 DATA ANGRY
370 DATA HAPPY
380 DATA JEALOUS
390 DATA BORED
400 DATA FED UP
410 DATA EMBARRASSED
420 DATA FURIOUS
430 DATA BORED
440 DATA FURIOUS
450 DATA HAPPY
460 DATA SAD
470 DATA EMBARRASSED
480 DATA JEALOUS
490 DATA BORED
500 DATA HAPPY
510 DATA ANGRY
520 DATA SAD
530 DATA FED UP
540 DATA EMBARRASSED
550 DATA JEALOUS
560 DATA ANGRY
570 DATA FURIOUS
580 DATA ANGRY
590 DATA FED UP
600 DATA FURIOUS
610 DATA JEALOUS
620 DATA SAD
```

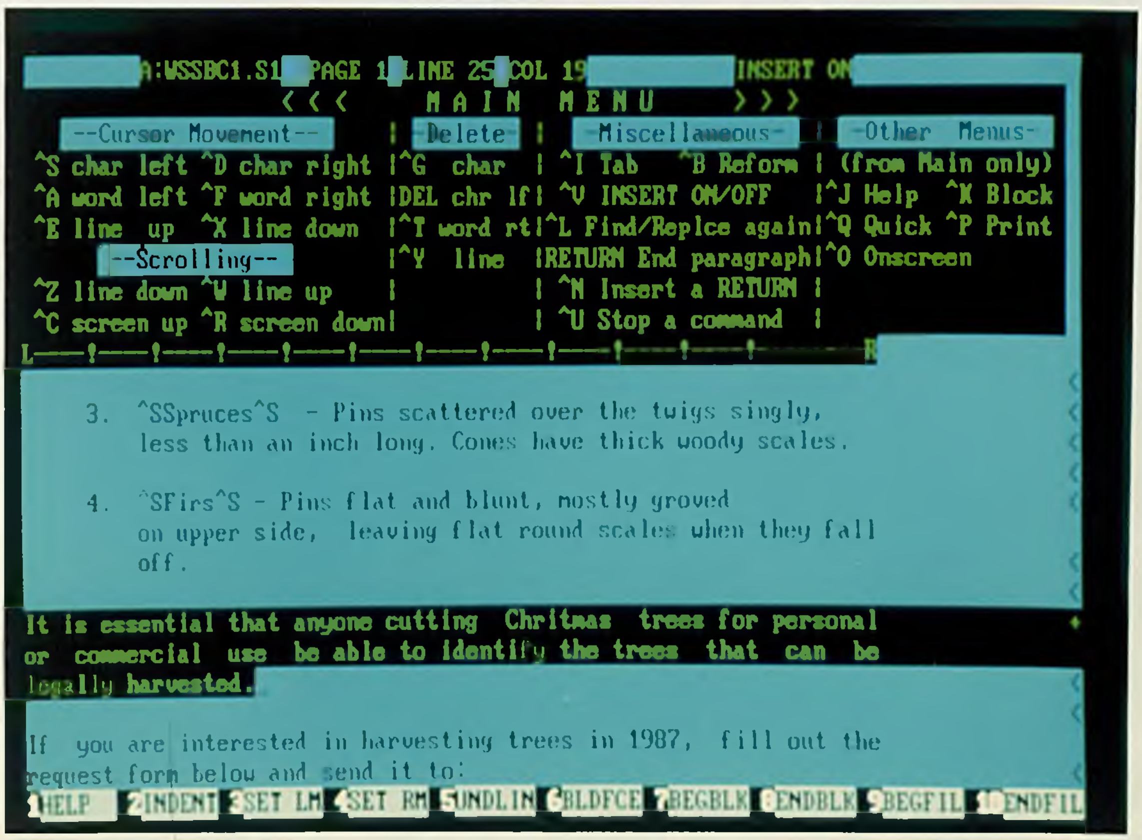
Appendix 7.1 Photographs of Screens: Word Perfect



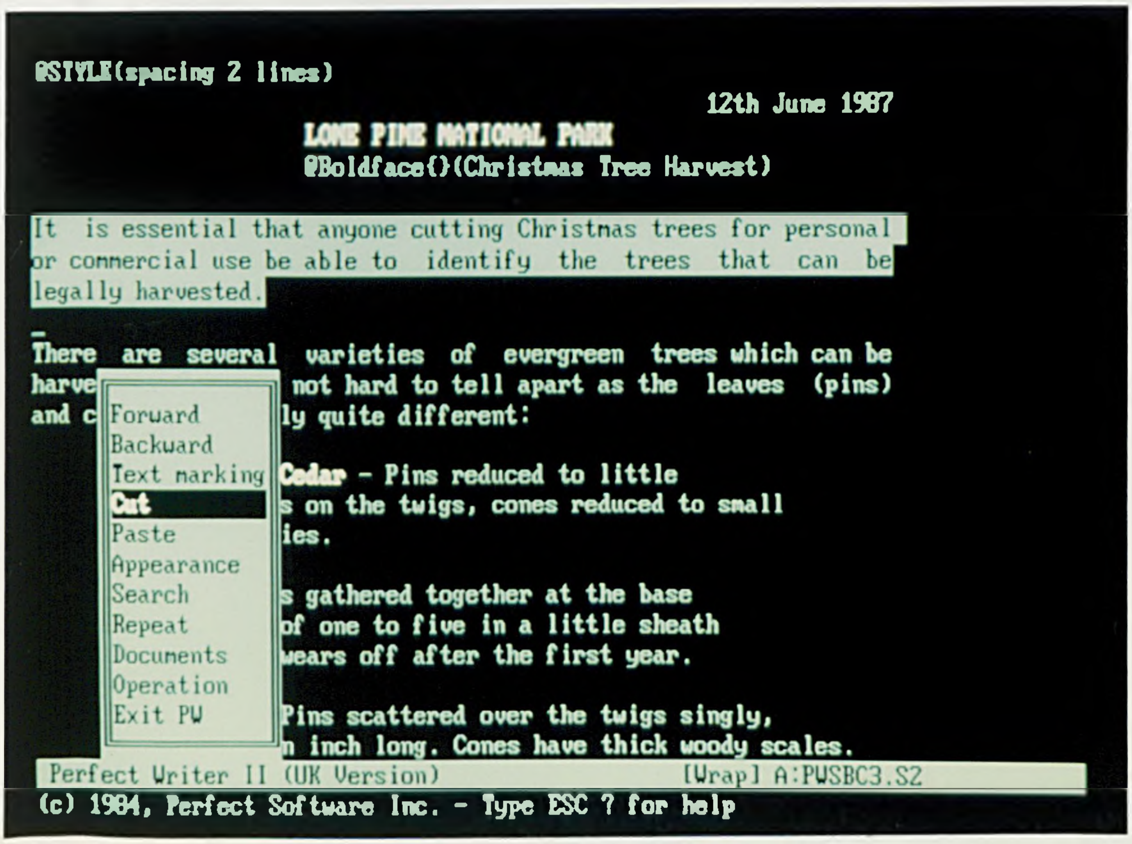


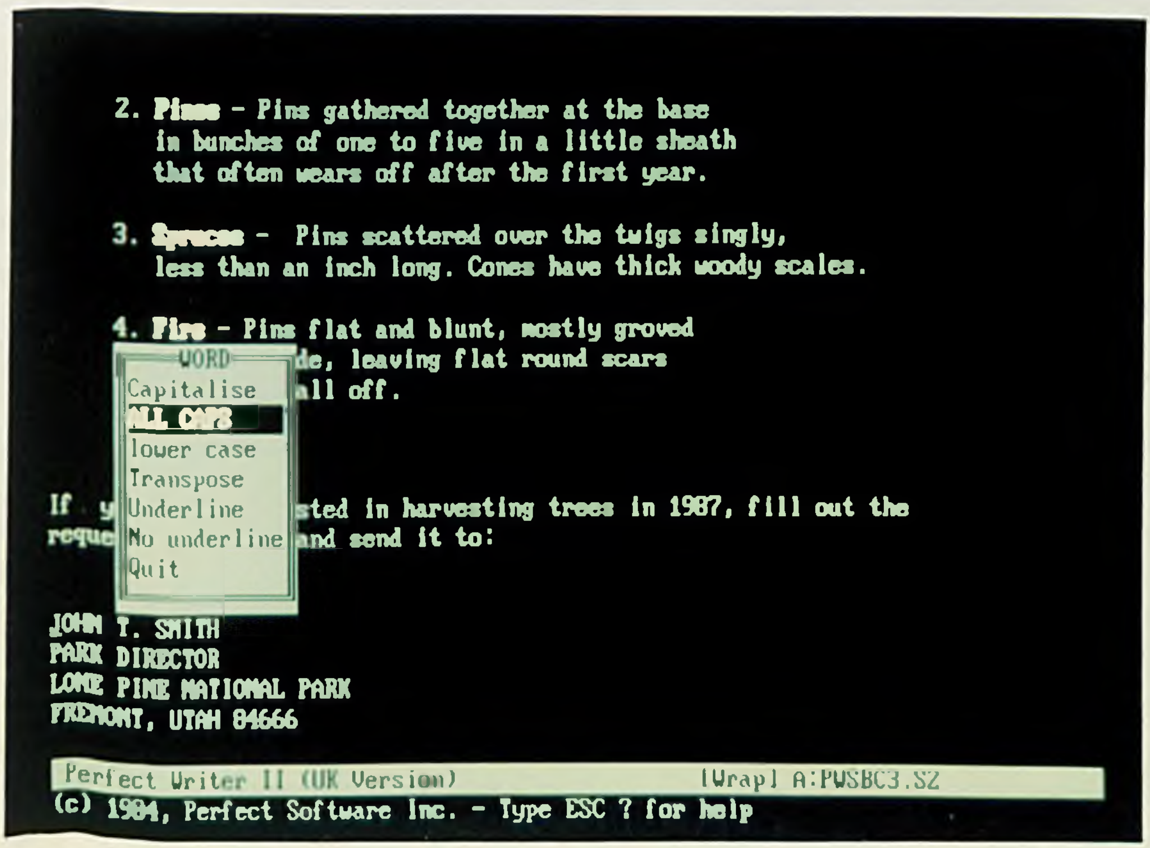
Photographs of Screens: Word Star





Photographs of Screens: Perfect Writer





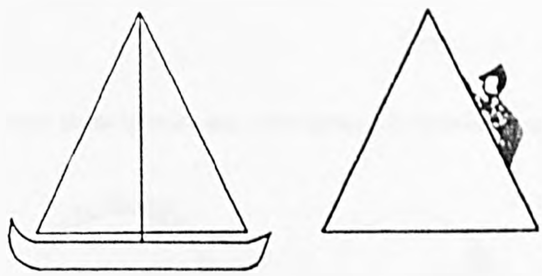
Appendix 7.2 Battery of Personal Styles Tests

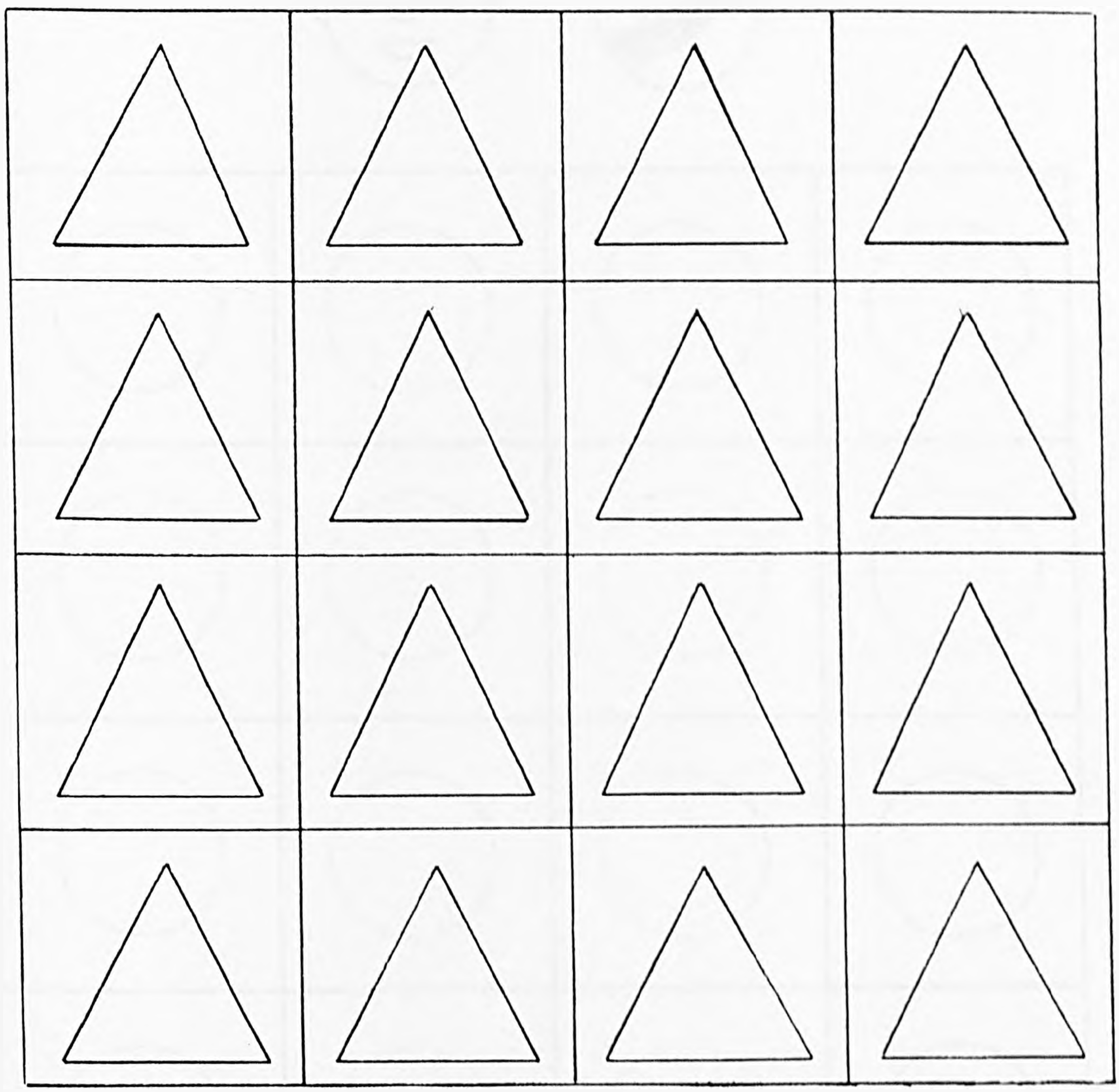
Divergent Test

Test DA3

Name ....

Complete as many as possible of these triangles to make them into a picture of a familiar thing (e.g.)



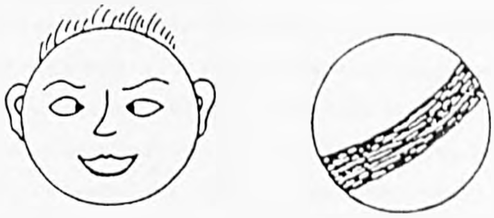


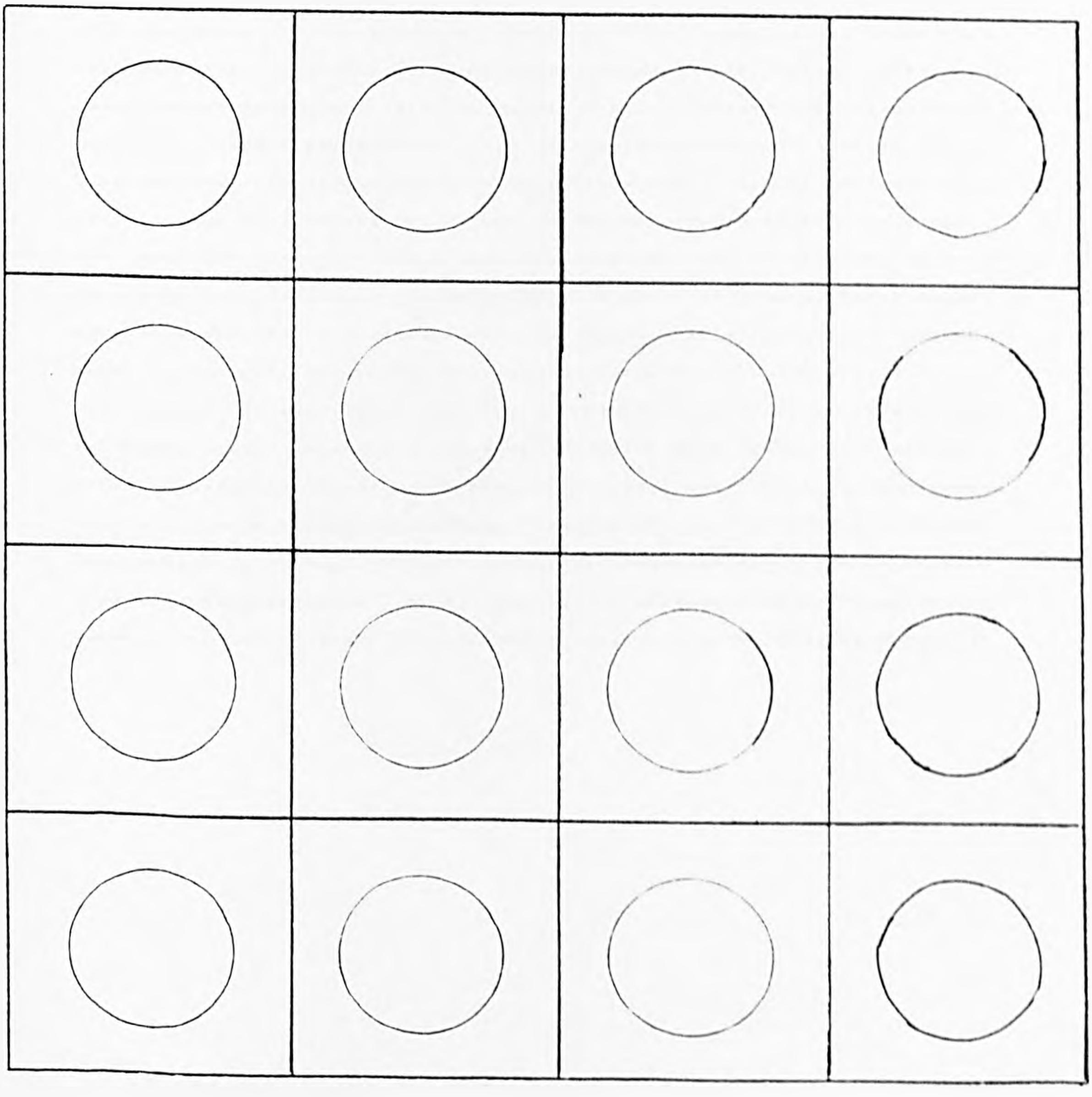
Divergent Test

Test DB3

Name ....

Complete as many as possible of these circles to make them into pictures of familiar things (e.g.)





Divergent Test

TEST DB4 NAME ....

Write down as many things as possible that SUGAR and SALT have in common

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

Divergent Test

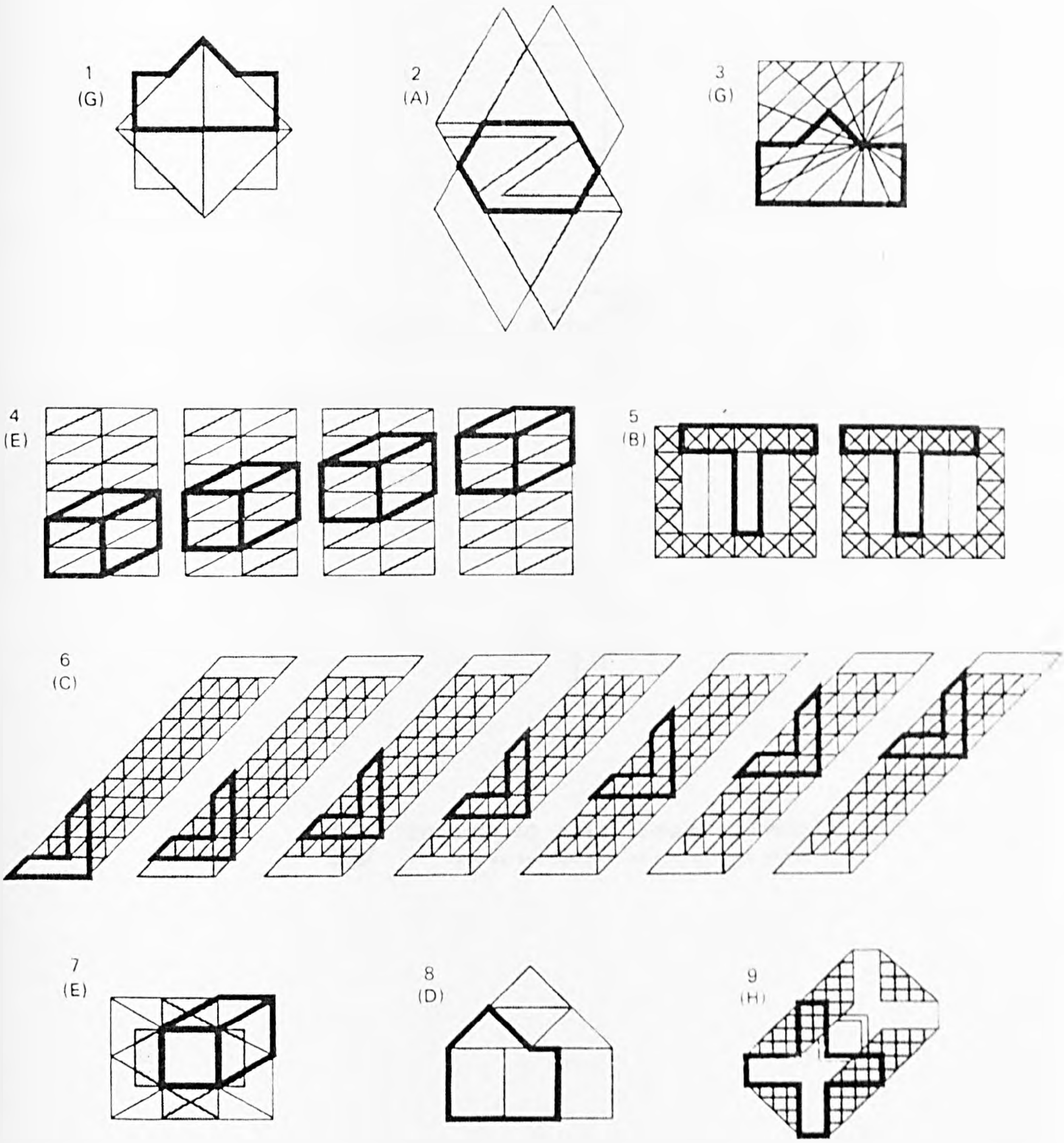
TEST DA4 NAME ....

Write down as may things as possible that MILK and MEAT have in common

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

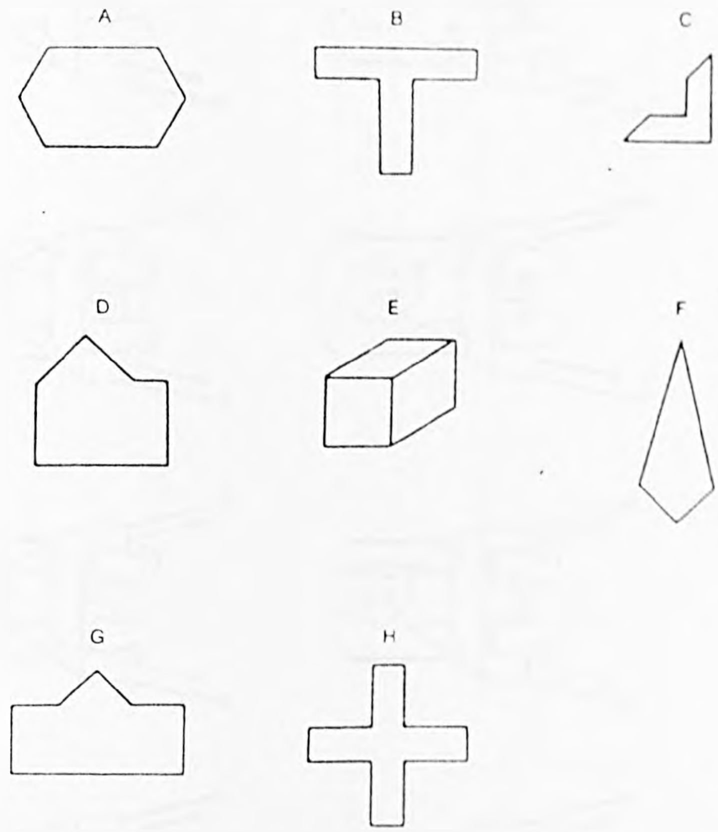
Examples of Items from the GEFT

GEFT-SCORING KEY SECOND SECTION



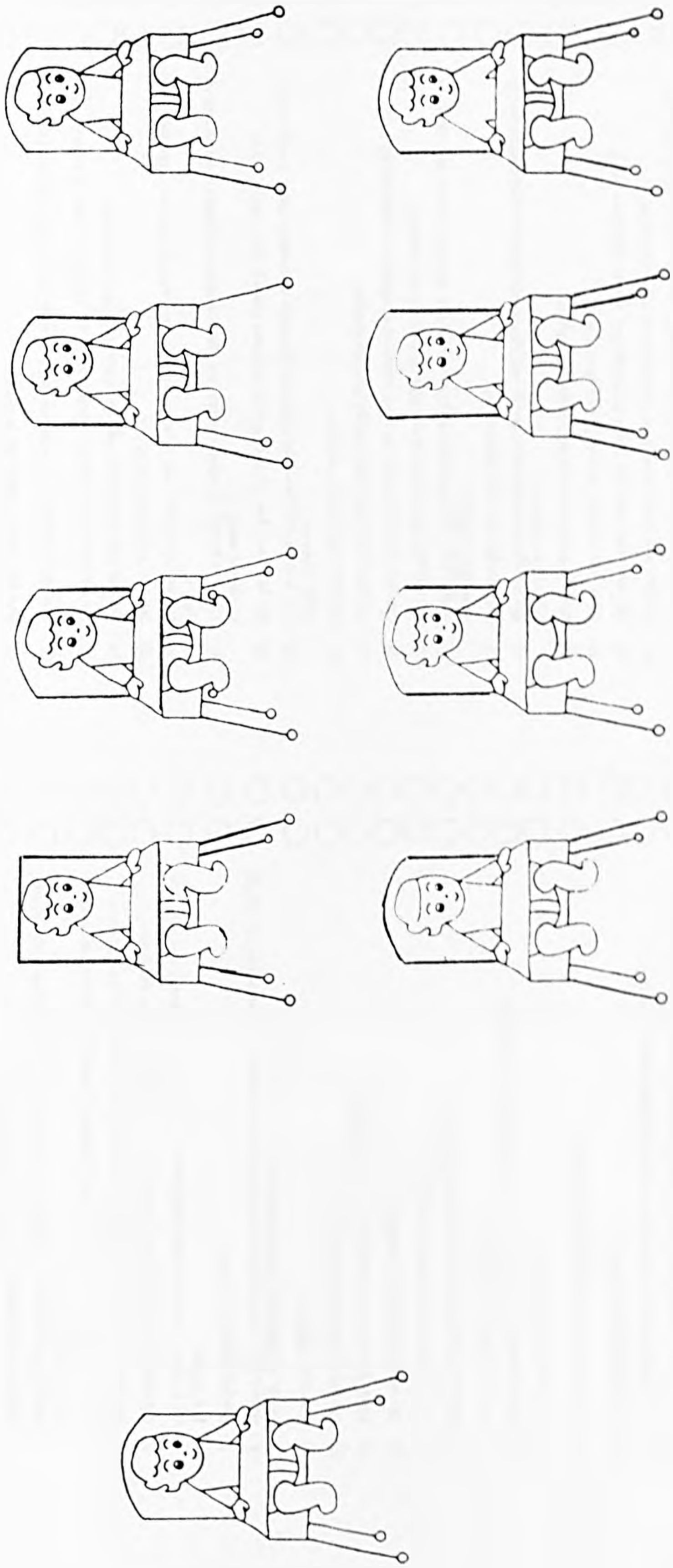
Items from the EFT and GEFT

SIMPLE FORMS



CONSULTING PSYCHOLOGISTS PRESS
577 College Avenue Palo Alto California 94306

One Item from the MFFT



Eysenck Personality Inventory

E ◯ N ◯ L ◯

FORM B

YES NO

1. Do you like plenty of excitement and bustle around you?
2. Have you often got a restless feeling that you want something but do not know what?
3. Do you nearly always have a "ready answer" when people talk to you?
4. Do you sometimes feel happy, sometimes sad, without any real reason?
5. Do you usually stay in the background at parties and "get-togethers"?
6. As a child, did you always do as you were told immediately and without grumbling?
7. Do you sometimes sulk?
8. When you are drawn into a quarrel, do you prefer to "have it out" to being silent, hoping things will blow over?
9. Are you moody?
10. Do you like mixing with people?
11. Have you often lost sleep over your worries?
12. Do you sometimes get cross?
13. Would you call yourself happy-go-lucky?
14. Do you often make up your mind too late?
15. Do you like working alone?
16. Have you often felt listless and tired for no good reason?
17. Are you rather lively?
18. Do you sometimes laugh at a dirty joke?
19. Do you often feel "fed-up"?
20. Do you feel uncomfortable in anything but everyday clothes?
21. Does your mind often wander when you are trying to attend closely to something?
22. Can you put your thoughts into words quickly?
23. Are you often "lost in thought"?
24. Are you completely free from prejudices of any kind?
25. Do you like practical jokes?
26. Do you often think of your past?
27. Do you very much like good food?
28. When you get annoyed, do you need someone friendly to talk to about it?
29. Do you mind selling things or asking people for money for some good cause?
30. Do you sometimes boast a little?
31. Are you touchy about some things?
32. Would you rather be at home on your own than go to a boring party?
33. Do you sometimes get so restless that you cannot sit long in a chair?
34. Do you like planning things carefully, well ahead of time?
35. Do you have dizzy turns?
36. Do you *always* answer a personal letter as soon as you can after you have read it?
37. Can you usually do things better by figuring them out alone than by talking to others about it?
38. Do you ever get short of breath without having done heavy work?
39. Are you an easy-going person, not generally bothered about having everything "just-so"?
40. Do you suffer from "nerves"?
41. Would you rather plan things than do things?
42. Do you sometimes put off until tomorrow what you ought to do today?
43. Do you get nervous in places like lifts, trains or tunnels?
44. When you make new friends, is it usually *you* who makes the first move, or does the inviting?
45. Do you get very bad headaches?
46. Do you generally feel that things will sort themselves out and come right in the end somehow?
47. Do you find it hard to fall asleep at bedtime?
48. Have you sometimes told lies in your life?
49. Do you sometimes say the first thing that comes into your head?
50. Do you worry too long after an embarrassing experience?
51. Do you usually keep "yourself to yourself" except with very close friends?
52. Do you often get into a jam because you do things without thinking?
53. Do you like cracking jokes and telling funny stories to your friends?
54. Would you rather win than lose a game?
55. Do you often feel self-conscious when you are with superiors?
56. When the odds are against you, do you still usually think it worth taking a chance?
57. Do you often get "butterflies in your tummy" before an important occasion?

PLEASE CHECK TO SEE THAT YOU HAVE ANSWERED ALL THE QUESTIONS

Styles of Learning Questionnaire (supplied by N.J. Entwistle)

In this section we would like you to show whether you agree or disagree with each of the statements listed below. We are concerned here with your approaches to studying in general. If your answer would be different for different subjects, however, you should reply in relation to your main course or subject.

Please circle the number beside each statement which best conforms with your view.

4 (✓✓) means Definitely agree

3 (✓) means Agree with reservations

1 (x) means Disagree with reservations

0 (xx) means Definitely disagree

2 (?) is only to be used if the item doesn't apply to you or if you find it impossible to give a definite answer.

		✓✓	✓	x	xx	?	
1.	I find it difficult to organise my study time effectively.	4	3	1	0	2	16
2.	I try to relate ideas in one subject to those in others, whenever possible.	4	3	1	0	2	17
3.	Although I have a fairly good general idea of many things, my knowledge of the details is rather weak.	4	3	1	0	2	18
4.	I enjoy competition: I find it stimulating.	4	3	1	0	2	19
5.	I usually set out to understand thoroughly the meaning of what I am asked to read.	4	3	1	0	2	20
6.	Ideas in books often set me off on long chains of thought of my own, only tenuously related to what I was reading.	4	3	1	0	2	21
7.	I chose my present courses mainly to give me a chance of a really good job afterwards.	4	3	1	0	2	22
8.	Continuing my education was something which happened to me, rather than something I really wanted for myself.	4	3	1	0	2	23

	√√	√	x	x	?	
9. I like to be told precisely what to do in essays or other assignments.	4	3	1	0	2	24
10. I often find myself questioning things that I hear in lectures or read in books.	4	3	1	0	2	25
11. I generally prefer to tackle each part of a topic or problem in order, working out one at a time.	4	3	1	0	2	26
12. The continual pressure of work—assignments, deadlines and competition—often makes me tense and depressed.	4	3	1	0	2	27
13. I find it difficult to "switch tracks" when working on a problem: I prefer to follow each line of thought as far as it will go.	4	3	1	0	2	28
14. My habit of putting off work leaves me with far too much to do at the end of term.	4	3	1	0	2	29
15. It's important to me to do really well in the courses here.	4	3	1	0	2	30
16. Lecturers seem to delight in making the simple truth unnecessarily complicated.	4	3	1	0	2	31
17. Distractions make it difficult for me to do much effective work in the evenings.	4	3	1	0	2	32
18. When I'm doing a piece of work, I try to bear in mind exactly what that particular lecturer seems to want.	4	3	1	0	2	33
19. I usually don't have time to think about the implications of what I have read.	4	3	1	0	2	34
20. Lecturers sometimes give indications of what is likely to come up in exams, so I look out for what may be hints.	4	3	1	0	2	35
21. In trying to understand a puzzling idea, I let my imagination wander freely to begin with, even if I don't seem to be much nearer a solution.	4	3	1	0	2	36
22. My main reason for being here is that it will help me to get a better job.	4	3	1	0	2	37
23. Often I find myself wondering whether the work I am doing here is really worthwhile.	4	3	1	0	2	38

	√√	√	x	x	?	
24. I generally put a lot of effort into trying to understand things which initially seem difficult.	4	3	1	0	2	39
25. I prefer courses to be clearly structured and highly organised.	4	3	1	0	2	40
26. A poor first answer in an exam makes me panic.	4	3	1	0	2	41
27. I prefer to follow well tried approaches to problems rather than anything too adventurous.	4	3	1	0	2	42
28. I'm rather slow at starting work in the evenings.	4	3	1	0	2	43
29. In trying to understand new ideas, I often try to relate them to real life situations to which they might apply.	4	3	1	0	2	44
30. When I'm reading I try to memorise important facts which may come in useful later.	4	3	1	0	2	45
31. I like to play around with ideas of my own even if they don't get me very far.	4	3	1	0	2	46
32. I generally choose courses more from the way they fit in with career plans than from my own interests.	4	3	1	0	2	47
33. I am usually cautious in drawing conclusions unless they are well supported by evidence.	4	3	1	0	2	48
34. When I'm tackling a new topic, I often ask myself questions about it which the new information should answer.	4	3	1	0	2	49
35. I suppose I am more interested in the qualifications I'll get than in the courses I'm taking.	4	3	1	0	2	50
36. Often I find I have to read things without having a chance to really understand them.	4	3	1	0	2	51
37. If conditions aren't right for me to study, I generally manage to do something to change them.	4	3	1	0	2	52
38. In reporting practical work, I like to try to work out several alternative ways of interpreting the findings.	4	3	1	0	2	53

		✓✓	✓	x	xx	?	
39.	My main reason for being here is so that I can learn more about the subjects which really interest me.	4	3	1	0	2	54
40.	In trying to understand new topics, I often explain them to myself in ways that other people don't seem to follow.	4	3	1	0	2	55
41.	I find I have to concentrate on memorising a good deal of what we have to learn.	4	3	1	0	2	56
42.	It is important to me to do things better than my friends, if I possibly can.	4	3	1	0	2	57
43.	I find it better to start straight away with the details of a new topic and build up an overall picture in that way	4	3	1	0	2	58
44.	Often when I'm reading books, the ideas produce vivid images which sometimes take on a life of their own.	4	3	1	0	2	59
45.	One way or another I manage to get hold of the books I need for studying.	4	3	1	0	2	60
46.	I often get criticised for introducing irrelevant material into my essays or tutorials.	4	3	1	0	2	61
47.	I find that studying academic topics can often be really exciting and gripping.	4	3	1	0	2	62
48.	The best way for me to understand what technical terms mean is to remember the text-book definitions.	4	3	1	0	2	63
49.	I certainly want to pass the next set of exams, but it doesn't really matter if I only just scrape through.	4	3	1	0	2	64
50.	I need to read around a subject pretty widely before I'm ready to put my ideas down on paper.	4	3	1	0	2	65
51.	Although I generally remember facts and details, I find it difficult to fit them together into an overall picture.	4	3	1	0	2	66
52.	I tend to read very little beyond what's required for completing assignments.	4	3	1	0	2	67
53.	Having to speak in tutorials is quite an ordeal for me.	4	3	1	0	2	68

	✓✓	✓	x	xx	?	
54. Puzzles or problems fascinate me, particularly where you have to work through the material to reach a logical conclusion.	4	3	1	0	2	69
55. I spend a good deal of my spare time in finding out more about interesting topics which have been discussed in classes.	4	3	1	0	2	70
56. I find it helpful to 'map out' a new topic for myself by seeing how the ideas fit together.	4	3	1	0	2	71
57. I seem to be a bit too ready to jump to conclusions without waiting for all the evidence.	4	3	1	0	2	72
58. I hate admitting defeat, even in trivial matters.	4	3	1	0	2	73
59. I think it is important to look at problems rationally and logically without making intuitive jumps.	4	3	1	0	2	74
60. I find I tend to remember things best if I concentrate on the order in which the lecturer presented them.	4	3	1	0	2	75
61. When I'm reading an article or research report, I generally examine the evidence carefully to decide whether the conclusion is justified.	4	3	1	0	2	76
62. Tutors seem to want me to be more adventurous in making use of my own ideas.	4	3	1	0	2	77
63. When I look back, I sometimes wonder why I ever decided to come here.	4	3	1	0	2	78
64. I find academic topics so interesting, I should like to continue with them after I finish this course.	4	3	1	0	2	79
					1	80

Appendix 7.3 Norms for "Personal" Styles Tests

Norms for:

	EFT($\bar{x}$)	MFFT($\bar{x}$)	Inacc.	DA3	DB3	DA4	DB4	
				Divergent Tests				
S1	14.15	52.5	2	6	8	5	10	(29)
S2	42.78	51.34	3	3	5.5	4	7	(19.5)
S3	22.12	19.83	7	5	3	6.5	7	(21.5)
S4	18.8	63.28	5	4	9	5	8	(26)
S5	59.55	63.05	4	5	9.5	7	7	(28.5)
S6	14.96	48.94	4	4	4	6	7	(21)
S7	14.34	37.6	6	4	6	6	7	(23)
S8	20.8	49.53	6	6	9	8	7	(30)
S9	11.08	53.27	2	3	8.5	8	8	(27.5)
S10	13.66	21.29	4	3	5	7	7	(22)
S11	27.02	21.93	6	0	5	3	6	(14)
S12	10.32	17.4	5	4	9	6	4	(23)
S13	14.61	27.2	7	4	9	6	5	(24)
S14	22.77	32.0	4	7	7	10	7	(31)
S15	36.4							
S16	15.32							
S17	25.74							
S18	8.93							
S19	9.98							
S20	15.14							
S21	24.75							
S22	20.17							
S23	13.56							
S24	10.86							
S25	19.9							
S26	10.65							
S27	32.96							
S28	14.62							
S29	12.16							
S30	7.41							
S31	43.37							
S32	10.76							
S33	17.23							

Appendix 8.1 Materials for Study 4

Lone Pine National Park
Chritmas Tree Harvest delete word
add "s"

There are several varieties of evergreen trees which can be harvested. They are not not hard to tell apart as the leaves (needles) and cones are usually quite different:

insert space

2. Pines - Needles gathered together at the base in bunches of one to five in alittle sheath that often wears off after the first year.

3. Spruces - Needles gathered together at the base in bunches of one to five in a little sheath. delete sentence
Needles scattered over the twigs singly less than an inch long. Cones have thick woody scales. add comma

capital "N"

4. Firs - needles flat and blunt, mostly groved on upper side, leaveing flat roundscars
close up
when they fall off. delete letter insert space

delete spaces It is essential that anyone close up

cutting Christmas trees for personal or commercial be able to identify the trees that can be legally harvested.

delete spaces If you are interested in harvesting treees in 1987, fill out the request form below and send it to:

insert blank line delete letter

John T. Smith
Park Director
Lone Pine National Park
Fremont, Utah 84666

Centre text

Lone Pine National Park ← Underline & capitals
Christmas Tree Harvest ← bold face

There are several varieties of evergreen trees which can be harvested. They are not hard to tell apart as the leaves (needles) and cones are usually quite different:

Insert paragraph 1. (below)

2. Pines - Needles gathered together at the base in bunches of one to five in a little sheath that often wears off after the first year.

Underline

3. Spruces - Needles scattered over the twigs singly, less than an inch long. Cones have thick woody scales.

4. Firs - Needles flat and blunt, mostly groved on upper side, leaving flat round scars when they fall off.

It is essential that anyone cutting Christmas trees for personal or commercial use be able to identify the trees that can be legally harvested.

If you are interested in harvesting trees in 1987, fill out the request form below and send it to:

John T. Smith
Park Director
Lone Pine National Park

all capitals

Paragraph 1. Juniper and Cedar - Needles reduced to little green scales on the twigs, cones reduced to small bluish berries

Insert date and flush right

Lone Pine National Park

It is essential ...

Christmas Tree Harvest

There are several varieties of evergreen trees which can be harvested. They are not hard to tell apart as the leaves (needles) and cones are usually quite different:

Replace "needles" with "pins"

1. Juniper and Cedar - Needles reduced to little green scales on the twigs; cones reduced to small bluish berries.

Replace "Needles" with "Pins"

2. Pines - Needles gathered together at the base in bunches of one to five in a little sheath that often wears off after the first year.

3. Spruces - Needles scattered over the twigs singly, less than an inch long. Cones have thick woody scales.

4. Firs - Needles flat and blunt, mostly groved on upper side, leaving flat round scars when they fall off.

Move parag. to top of page

It is essential that anyone cutting Christmas trees for personal or commercial use be able to identify the trees that can be legally harvested.

If you are interested in harvesting trees in 1987, fill out the request form below and send it to:

Turn justification off

JOHN T. SMITH
PARK DIRECTOR
LONE PINE NATIONAL PARK

Number Page

12 June 1987

Lone Pine National Park

Christmas Tree Harvest

It is essential that anyone cutting Christmas trees for personal or commercial use be able to identify the trees that can be legally harvested.

There are several varieties of evergreen trees which can be harvested. They are not hard to tell apart as the leaves (pins) and cones are usually quite different:

1. Juniper and Cedar - Pins reduced to little green scales on the twigs, cones reduced to small bluish berries.

2. Pines - Pins gathered together at the base in bunches of one to five in a little sheath that often wears off after the first year.

3. Spruces - Pins scattered over the twigs singly, less than an inch long. Cones have thick woody scales.

4. Firs - Pins flat and blunt, mostly groved on upper side, leaving flat round scars when they fall off.

If you are interested in harvesting trees in 1987, fill out the request form below and send it to:

JOHN T. SMITH

PARK DIRECTOR

LONE PINE NATIONAL PARK

Other Tasks

1. Save the file you have created and call it "myfile" (use your name)

2. Change the margines so that the left margin is 5 and the right margin is 70 (those using Word Perfect) and left margin is 5 and the right margin is 60 (those using Word Star and Perfect Writer)

3. Reformat the text

4. Save the file

5. Rename the file "endfile"

6. Print "endfile"

Appendix 8.2 Questionnaires and Visual Analogue Scales

EQ

Consider the two ends of the line as representing your most extreme feelings. Put a mark on the line to show how you feel right now.

NOT AT ALL TENSE	________________________	TENSE
NOT AT ALL BORED	________________________	BORED
NOT AT ALL TIRED	________________________	TIRED
NOT AT ALL SAD	________________________	SAD

N.B. not to scale

Visual Analogue Scales for Difficulty

Each line represents a dimension of difficulty for categories of word processing skills. Please rate each category on how difficult you found it to perform these skills, taking the two ends of the line as your most extreme feelings.

Moving cursor around screen

not so difficult ———————————————— very difficult

Editing text

not so difficult ———————————————— very difficult

Highlighting text

not so difficult ———————————————— very difficult

Formating text

not so difficult ———————————————— very difficult

Printing text

not so difficult ———————————————— very difficult

File handling

not so difficult ———————————————— very difficult

N.B. not to scale

Questionnaires to Assess Motivation and Attitudes to Computers

Questionnaire A1

Please rate the following statements on a scale of -3 to +3 where a minus number indicates disagreement with the statements below and a positive number indicates agreement. Zero indicates undecided or indifferent.

1. I believe that computers are essential to society. -3 -2 -1 0 +1 +2 +3
2. I believe that computers are helpful -3 -2 -1 0 +1 +2 +3
3. I believe that computers create unemployment. -3 -2 -1 0 +1 +2 +3
4. I believe that computers will create disharmony in society. -3 -2 -1 0 +1 +2 +3
5. I believe that computers are anti-social. -3 -2 -1 0 +1 +2 +3
6. I believe that computers are a health risk. -3 -2 -1 0 +1 +2 +3
7. I am fightened or nervous of computers. -3 -2 -1 0 +1 +2 +3

Questionnaire M1

Please rate the following statements on a scale of -3 to +3 where a minus number indicates disagreement with the statements below and a positive number indicates agreement. Zero indicates undecided or indifferent.

1. I think I would enjoy working with computers. -3 -2 -1 0 +1 +2 +3
2. I would like to do a computer course. -3 -2 -1 0 +1 +2 +3
3. I would use computers at work or in my studies only if I was forced to do so. -3 -2 -1 0 +1 +2 +3
4. I enjoy playing computer games. -3 -2 -1 0 +1 +2 +3
5. If I had access to a computer I would use it frequently. -3 -2 -1 0 +1 +2 +3
6. If I had the money I would buy a home computer. -3 -2 -1 0 +1 +2 +3

Appendix 8.3

Accumulative Increases in Ratings for Tiredness

Subject 1	WS	WP	PW
Day 1	4.5 (31.5)*	8.5 (59.0)	0.0 (72.0)
Day 2	0.5 (59.0)	23.0 (11.5)	26.0 (46.0)
Day 3	1.5 (41.5)	.5 (29.5)	0.0 (57.0)

Subject 2	WS	WP	PW
Day 1	48.0 (7.5)	0.0 (47.5)	4.0 (17.0)
Day 2	56.0 (5.5)	2.0 (22.5)	36.0 (22.0)
Day 3	15.0 (20.5)	0.0 (55.0)	32.0 (18.5)

Subject 3	WS	WP	PW
Day 1	17.0 (49.0)	11.0 (62.0)	16.5 (39.5)
Day 2	28.0 (50.0)	2.5 (47.5)	67.0 (22.0)
Day 3	1.0 (52.0)	11.5 (47.0)	0.0 (57.0)

Acumulative Increases in Ratings for Sadness

Subject 1	WS	WP	PW
Day 1	8.0 (2.0)	0.0 (33.0)	5.5 (7.0)
Day 2	0.0 (6.0)	0.0 (12.0)	0.0 (78.0)
Day 3	0.0 (2.5)	0.0 (6.0)	2.0 (42.0)

Subject 2	WS	WP	PW
Day 1	29.5 (6.5)	0.0 (27.5)	0.0 (7.0)
Day 2	17.5 (5.0)	3.0 (22.5)	0.0 (4.0)
Day 3	1.0 (7.0)	0.0 (28.5)	0.0 (9.5)

Subject 3	WS	WP	PW
Day 1	0.0 (50.0)	33.0 (40.5)	0.0 (51.0)
Day 2	0.0 (50.0)	5.5 (46.0)	0.0 (67.0)
Day 3	5.0 (51.0)	45.0 (46.0)	33.0 (34.0)

WS=Word Star
WP=Word Perfect
PW=Perfect Writer

*Values in parentheses refer to the starting value on the VAS

Appendix 9.1 Frequency Count of Subject 1's Misunderstandings

	C1	C2	C3	Total
Basic editing problems	11	6	8	25 (1)
Inccrrect command used	12	0	5	17 (2)
Locating a command	2	4	3	9 (9)
Did not read screen	12	2	1	15 (3)
How to carry out a command	3	4	4	11 (6=)
Difficulty with menus	8	0	6	14 (4)
Incorrect reasoning	7	2	3	12 (5)
Don't know function of a command	7	2	2	11 (6=)
Misunderstands function	7	1	2	10 (8)
Wrong cursor position	5	1	2	8 (10)
Changes not understood	1	0	1	2 (17)
Misreading manual	0	4	1	5 (12=)
False assumption of error	3	0	1	4 (14)
Typing in at beginning	2	1	2	5 (12=)
Typeover v's deletion	4	1	1	6 (11)
Inconsistent procedures	0	2	1	3 (15=)
Does not realise a change	3	0	0	3 (15=)
Incorrect spelling/case	1	0	0	1 (18)
Total	88	30	43	

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

Appendix 9.2 Frequency Count of Subject 2's Misunderstandings

	C1	C2	C3	Total
Basic editing problems	4	2	2	8 (2=)
Incorrect command used	6	1	1	8 (2=)
Locating a command	5	2	5	12 (1)
Did not read screen	1	2	3	6 (6=)
How to carry out a command	2	1	2	5 (8=)
Difficulty with menus	3	0	3	6 (6=)
Incorrect reasoning	2	4	2	8 (2=)
Don't know function of a command	7	2	2	11 (6=)
Misunderstands a function	2	0	0	2 (16=)
Wrong cursor position	1	2	1	4 (10)
Changes not understood	1	2	0	3 (11=)
Misreading manual	1	3	1	5 (8=)
False assumption of error	2	0	1	3 (11=)
Typing in at beginning	0	1	1	2 (14=)
Typeover v's deletion	0	3	0	3 (11=)
Inconsistent procedures	1	0	0	1 (17=)
Does not realise a change	0	2	0	2 (14=)
Incorrect spelling/case	1	0	0	1 (17=)
Total	36	28	23	

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

Appendix 9.3 Frequency Count of Subject 3's Misunderstandings

	C1	C2	C3	Total
Basic editing problems	8	3	12	23 (1)
Incorrect command used	6	1	4	11 (3=)
Locating a command	5	2	4	11 (3=)
Did not read screen	5	3	1	9 (6)
How to carry out a command	0	6	6	12 (2)
Difficulty with menus	1	0	5	6 (8)
Incorrect reasoning	0	3	1	4 (10)
Don't know function of a command	3	2	5	10 (5)
Misunderstands a function	0	1	4	5 (9)
Wrong cursor position	2	1	5	8 (7)
Changes not understood	1	0	1	2 (14=)
Misreading manual	1	1	0	2 (14=)
False assumption of error	0	2	1	3 (11=)
Typing in at beginning	1	0	1	2 (14=)
Typeover v's deletion	0	0	0	0 (18)
Inconsistent procedures	1	1	1	3 (11=)
Does not realise a change	1	0	0	1 (17)
Incorrect spelling/case	1	0	2	3 (11=)
Total	36	26	53	

Key: C1 = Word Star
C2 = Word Perfect
C3 = Perfect Writer

Appendix 10.1 Visual Analogue Scales

Pre-test Questionnaire

Name:
Age:

1. Have you ever written a computer program?
   If so please rate how proficient you are at computer programming by placing a mark on the line below.

not at all proficient	________________________________	extremely proficient

2. Have you ever used a word processor? If so please state the number of word processors you have used and their names.

   Please put a mark on the two lines below to show how proficient and how frequently you use a word processor.

not at all proficient	________________________________	extremely proficient
not at all often	________________________________	extremely often

3. Have you ever used a computer for statistical analyses? If so please put a mark on the two lines below to show how proficient and how frequently you use a computer for statistical analyses.

not at all proficient	________________________________	extremely proficient
not at all often	________________________________	extremely often

4. In general, would you describe yourself as:

not at all experienced	________________________________	extremely experienced

in the use of computers.

5. How much do you like computers?

not at all	________________________________	a great deal

N.B. not to scale

8. Name as many options as possible which can be found under the "typeface" command.

9. Are documents automatically page numbered?

10. How do you switch form "insert" mode to "overwrite" mode? What is the difference between them?

11. How do you name a document?

12. What is the sequence of commands that allow you to leave the document?

13. Name as many functions as you can which are <u>not</u> carried out within the document.

14. What is the sequence of commands for "underlining" text?

Post-test Questionnaire

1. How difficult did you find this word processor to use?

not at all difficult	________________________________	extremely difficult

2. How similar is this word processor to any other word processor that you have used?

not at all similar	________________________________	extremely similar

Please answer the following questions as fully as possible. If you cannot answer any of the questions, you may guess. Please indicate if your answer is a guess. If you cannot guess, go straight onto the next question.

3. Give a description of how the word processor works. For example, how is the information structured and how do you go about giving commands.

4. Into how many windows or sections is the screen divided and what are their functions?

5. What does the F3 key do?

6. How do you change the margins of your document?

7. What does the "design" command allow you to do?

Appendix 10.2 Instructions to Subject

This study involves learning to use a word processor called "Xchange". A short piece of text has already been typed. Your task is to make as many alterations to the text in 20 minutes and also to learn as much as possible about how the word processor works.

You may ask the experimenter for specific information about how to accomplish tasks.

The Test

1. Type in "xchange"

2. Select "Quill"

3. Load a document called "buffalo"

4. Do the following alterations:

a. Underline <u>Lone Pine National Park</u>

b. Bold BUFFALO PERMITS

c. Centre "1984"

d. Delete spaces in paragraph 1. last line

e. Add a "t" to lotery

f. Insert a blank line between "permit Lottery" and "1984"

g. Replace "bufalo" with "buffalo" using a global replace i.e., replacing all "bufalos" with one command

h. Change the left margin to 5

i. Justify (make straight) right margin

j. Change page length from 66 to 72 lines

k. Save document and call it "myfile"

l. Rename the document with your initials

m. Print the document

Appendix 10.3 Passage used in Study 5

Underline

Lone Pine National Park

BUFFALO PERMITS ← Boldface

General Season

1984 Centre text

Justify Right margin

A 3-day season, December 31-February 2, for the taking of bufalo is prescribed for the Lone Pine Game Reserve. A person with a valid bufalo permit per season. In addition, hunters are required to attend a Bufalo Orientation Seminar where light refreshments will be served. Replace "bufalo" with "buffalo"

delete spaces

add "t"

Permit Lotery

Insert blank line 1984

The Board of Big Game Control established seasons, permit numbers and permit lottery procedures for the current bufalo hunt. The lottery will be held at Lone Pine Visitor's Centre on December 25, 1984 at 7.00 a.m.

Change left margin to 5

Page length = 72 lines

Appendix 10.4 Supplementary Appendices

Study 2: Raw Data

	No. of Errors			No. of Trials			Last Trial Response Time(sec)			
S	COL	SHA	POS	COL	SHA	POS	COL	SHA	POS	EFT
1(m)	00	01	09	1	2	4	2.15	2.11	3.92	14.61
2(m)	24	21	27	5	5	5	0.76	0.80	4.19	09.85
3(f)	02	04	06	2	3	4	1.35	1.34	5.50	33.66
4(m)	06	21	06	4	5	4	1.61	1.63	4.64	-
5(f)	04	24	13	3	5	4	0.53	1.35	4.77	59.55
6(f)	15	05	29	5	3	5	1.63	0.85	3.07	22.12
7(f)	01	02	04	2	2	2	0.72	0.93	2.36	21.85
8(f)	16	01	15	4	2	5	1.50	1.09	2.89	-
9(f)	01	05	08	2	4	5	0.92	0.81	3.20	10.86
10(m)	09	08	25	4	5	5	2.10	1.47	6.80	14.96
11(m)	37	09	43	5	5	5	1.16	1.27	4.10	20.17
12(f)	08	00	02	5	1	2	0.85	1.38	3.15	14.34
13(m)	16	14	16	5	5	5	1.56	1.44	5.53	10.65
14(f)	09	10	14	4	5	5	1.64	1.82	3.06	27.02
15(f)	03	08	12	2	3	4	2.24	1.54	5.38	15.32
16(m)	05	08	40	3	3	5	3.89	4.97	4.19	42.78
17(m)	12	00	15	4	1	5	2.59	1.22	8.76	14.15
18(m)	15	01	01	5	2	2	1.93	1.75	6.56	11.08
19(m)	02	00	00	2	1	5	2.11	1.49	2.90	12.16
20(f)	04	05	03	2	3	2	1.31	2.45	4.80	07.91
21(f)	01	13	26	2	4	5	3.82	1.77	7.43	43.37
22(f)	08	13	08	3	4	4	1.00	1.26	4.03	20.80
23(m)	03	00	21	2	1	5	1.52	1.63	3.23	10.76
24(m)	07	00	23	4	1	5	1.27	1.12	2.82	17.23

Key: COL = colour
SHA = shape
POS = position

Appendix 10.4 Supplementary Appendices

Study 3: Subjects Scores after Log Transformations

Subjects	Baseline	Appropriate	Inappropriate
S1 (f)	2.32	2.37	2.35
S2 (f)	2.53	2.31	2.36
S3 (m)	2.42	2.38	2.54
S4 (m)	2.53	2.46	2.40
S5 (f)	2.43	2.19	2.39
S6 (f)	2.30	2.41	2.43
S7 (m)	2.48	2.49	2.52
S8 (m)	2.46	2.44	2.48
S9 (f)	2.35	2.38	2.42
S10 (m)	2.55	2.59	2.54
S11 (f)	2.37	2.35	2.45
S12 (f)	2.25	2.26	2.26
S13 (m)	2.42	2.52	2.52
S14 (m)	2.62	2.66	2.75
S15 (m)	2.43	2.32	2.36
S16 (f)	2.42	2.53	2.53
S17 (m)	2.27	2.38	2.26
S18 (f)	2.47	2.57	2.55
x	2.42	2.42	2.46
SD	0.01	0.12	0.11

Appendix 10.4 Supplementary Appendices

Study 5: Subjects Scores for on Pre- and Post Tests

	Pre-test Questions								Post-		Tasks	Compre
	1	2A	2B	2C	3A	3B	4	5	1	2	Comp	Scores
S1	00	01	69	72	59	66	59	14	63	25	15	17
S2	00	06	60	74	53	37	48	34	38	32	14	23
S3	09	03	68	68	77	76	49	87	41	57	16	24
S4	02	03	47	72	21	31	19	63	66	40	12	17
S5	85	09	90	92	91	94	94	24	33	60	16	32
S6	16	01	92	63	95	72	72	81	75	09	11	19.5
S7	00	01	49	49	61	60	29	53	53	26	09	05
S8	15	01	42	27	00	00	27	34	73	10	06	03
S9	00	00	00	00	00	00	03	50	83	00	15	20
S10	00	01	22	19	69	66	31	54	85	18	06	10
S11	19	03	57	62	75	79	38	53	57	70	10	14
S12	41	04	73	77	74	73	74	88	64	47	12	20
S13	00	01	39	46	20	17	09	44	49	16	15	10

Table to show frequency of question-asking

	Con	What?	What now?	How?	Why?	Do I?	Ind	Dir	Tot	HM
S1	10	03	04	13	06	30	02	56	68	00
S2	09	00	00	05	02	13	19	20	48	02
S3	12	01	00	05	05	34	00	45	57	01
S4	11	03	02	13	06	21	07	45	63	01
S5	00	00	00	02	00	00	01	02	03	09
S6	10	03	00	03	03	35	05	44	59	06
S7	06	00	02	11	01	09	04	23	33	02
S8	03	02	02	04	02	18	10	28	41	00
S9	44	04	07	27	02	28	01	68	73	00
S10	05	01	02	07	03	00	05	13	23	04
S11	10	03	02	11	03	08	07	27	44	00
S12	00	00	00	05	00	01	04	06	10	04
S13	07	01	02	21	06	25	00	55	62	00

KEY: Con = Confirmation
What? = What is x?
What now? = What do I do now?
Ind = Indirect
Tot = Total
Hm = Help menu