Personalising Interaction Using User Interface Skins

A thesis submitted for the degree of Doctor of Philosophy

Nick Fine

School of Computing, Information Systems and Mathematics

Brunel University

November 2008

Abstract

This research presents a method for personalising interaction using user interface skins, using personality as a design segment, enabling design for large anonymous user populations. User interface skins represent the audio-visual elements of the user interface, for example background colour or error tone, and can provide a means for the user to quickly and easily change their own user interfaces - and therefore the quality of their own interaction. However, there is limited research available as to how user interface skins can be best used to provide a more personalised and therefore more personally relevant user experience in comparison to a typical single user interface approach, designed for the average intended user.

In this research, three types of user interface element are investigated: colour, shape and meaning. The first two are thought to be operationalised by psychophysiological arousal. Meaning in user interface skins is investigated through a testing of the Similarity Attraction Hypothesis, to determine if people choose interface skins that exhibit the same high levels of certain personality traits that they themselves score highly in. Using an externalised methodology facilitated by the Internet, a client server experimental platform was developed to investigate four experimental hypotheses. The results of the hypothesis testing indicated support for the first experimental hypothesis, that personality can be determined from log files. The second hypothesis, that colour can be linked to personality, was supported by evidence indicating personality is a factor in preference for user interface colour. The third hypothesis, that shape can be linked to personality, was not supported. The fourth experimental hypothesis, investigating the Similarity Attraction Hypothesis, was supported with evidence from the Extroversion trait. However, for Neuroticism strong evidence for a Dissimilarity Attraction Effect was found. The overall experimental hypothesis that a more personalised user interaction can be produced using profiled user interface skins was supported through the evidence found in the experimental hypotheses, with the main exception of shape. Personality was found to be an HCI relevant design factor. The implications from these findings are discussed and reformulated experimental hypotheses presented for future work.

Declaration

The following publications have been produced, directly and indirectly, from the research undertaken in this investigation.

Fine, N., & Brinkman, W.-P. (2004). Avoiding Average: Recording Interaction Data to Design for Specific User Groups. In M. Rauterberg *Entertainment Computing – International Conference on Entertainment Computing (ICEC)* (p.398-401). Berlin: Springer

Fine, N., & Brinkman, W.-P. (2004). Informing Intelligent Environments: Creating Profiled User Interfaces. In E. van den Hoven, W. Ijsselsteijn, G. Kortuem, K. van Laerhoven, I. McClelland, E. Perik, N. Romero, and B. de Ruyter (ed.) *Adjunct Proceedings of European Symposium on Ambient Environments (EUSAI)*, (p. 15-17)

Brinkman, W.-P., & Fine, N. (2005). Towards Customized Emotional Design: An Explorative Study of User Personality and User Interface Skin Preferences. *Proceedings of Annual Conference of the European Association of Cognitive Ergonomics* (p. 107-114)

Fine, N. (2005). Personalising Interaction Using Profiled User Interface Skins. *Adjunct Proceedings from the British HCI Conference* 2005, p. 194-196.

Fine, N., & Brinkman, W.-P. (2006). Towards extracting personality trait data from interaction behaviour. In W.-P. Brinkman (ed.), *Proceedings of the 2006 Workshop on Computer Assisted Recording, Pre-Processing, and Analysis of User Interaction Data,* (p. 76-81). Morrisville, NC: Lulu.

Khan, I., Brinkman, W.-P., Fine, N. & Hierons, R. (2008) Measuring Personality from Keyboard and Mouse-related Interactive Behaviour. *Proceedings of European Conference on Cognitive Ergonomics (ECCE)* (p. 184-191)

Acknowledgements

The author would like to acknowledge the following people, without whom this thesis would not have been possible.

Dr. Willem-Paul Brinkman for his guidance and encouragement throughout. His patience is a virtue without which our working relationship and the entire experience would have been very different and certainly less enjoyable, and for that I am truly grateful.

Prof. Robert Macredie for his wisdom and insight, and especially "why this, why now?" that continues to influence every sentence I write, even this one.

My wife, Debbie, for her constant love and support in helping to achieve my ambitions, and for indulging my antisocial nocturnal working practices.

Andrew Evans for his deft programming skills in developing the ProSkin experiment platform and as a friend who graciously endured far too many of my frustrated moans.

Derry, Sissi and all the staff at Barneys, who provided a home and all the resources needed to write up this thesis and to whom I am forever grateful.

Finally to Prof. Harold Thimbleby for starting me out on this journey, and whose HCI teachings continue to guide me.

Table of Contents

Table of Figures

Tables

Chapter 1 - INTRODUCTION

This introductory chapter provides an overview of the research, detailing the motivations and research questions that have informed this thesis. First, the main research problem is presented. A potential solution is then described and this is formulated into a research question. The next section describes the approach taken to investigate the research question and serves to provide an overview of the thesis by referencing the relevant chapters. The chapter concludes with a visual guide to the thesis.

1.1 Introduction to the Research Problem

The user interface is the point at which human and computer interact and the multidisciplinary field of Human Computer Interaction (HCI) encompasses the study of user interfaces and the resulting interaction. This thesis investigates a method for personalising the user interface for different personality traits.

Personalisation of the user's computing experience is currently a topic of interest to academics, commercial developers and end users alike. Broadly speaking, personalisation can be sub-divided into two main categories: content and presentation. The personalisation of content is already a well established research area (e.g. Ferman et al (2002), Hoashi, Matsumoto and Inoue (2003), Encelle and Baptiste-Jessel (2007) and Chatzis, Doulamis and Varvarigou (2007)). However, the personalisation of presentation is still a relatively under-researched area. This is despite the fact that users currently have the tools to change their own interfaces, in the form of user interface skins (also known as themes). User interface skins typically comprise the aesthetic elements of the user interface and can include visual, aural and haptic elements. A brief example of a single application (Windows Media Player) with two alternative skins is illustrated in Figure 1:.

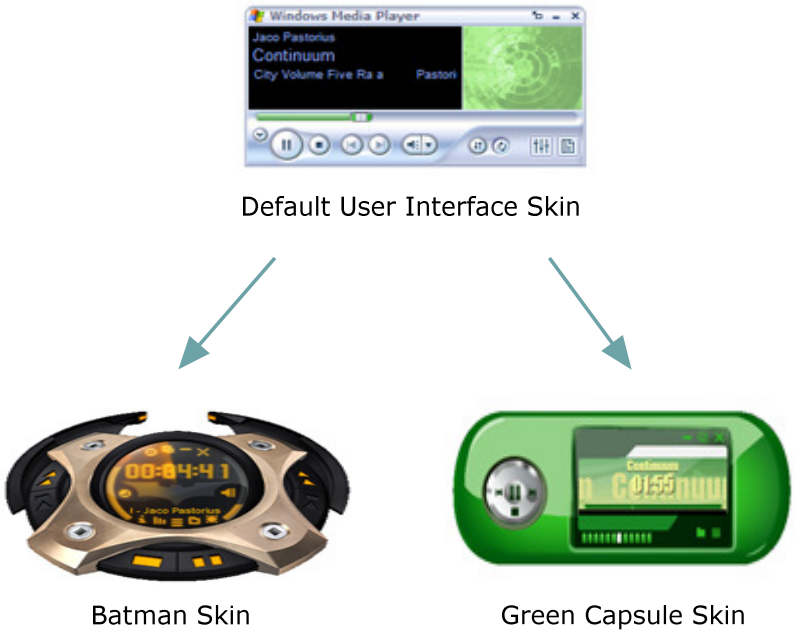


Figure 1:1 - Default user interface skin on an example application (Windows Media Player) with two alternative skins, illustrating a single application with different aesthetic effects enabled by user interface skins.

There are many examples of user-changeable user interfaces across a diverse number of applications: mobile phones, operating systems, Internet web browsers, media players, e-mail clients, web sites and software applications. Such applications feature user interface skins that the user themselves can change to something that they prefer. For example, the application may come with a default grey user interface skin, but the user can change that to either another colour, such as red, or to another theme or picture, such as a Star Wars theme. By utilising user interface skins to change the aesthetics of the user interface itself the user is able to personalise their experience. One of the assumptions made by this research is that a personally relevant, or personalised, interaction is preferable to a non-personalised, or generic user interface.

There is research into the nature of aesthetics as it relates to HCI (e.g. Tractinsky, Shoval-Katz and Ikar (2000), Norman (2002), Saari and Turpeinen (2004) and Tractinksy and Zmiri (2005)) that suggests that aesthetics plays an important role in HCI and can determine the quality of interaction. Blom and Monk (2003) present a theory of personalisation of appearance as an explanation for why users personalise their PCs and mobile phones. Their theory considers the user, system and context as

dispositions influencing the personalisation of appearance and cognitive, social and emotional effects resulting from the personalisation.

Logically it follows that changing the user interface also changes the quality of the interaction, so by empowering the user with the ability to change their own user interface, the user has also been empowered to change the quality of their own interaction – yet there is limited research in this area to inform how personalisation of presentation affects interaction. It is the aim of this research to investigate ways in which the user interface can be changed aesthetically to provide a personalised experience.

Another key motivation for this research is a desire to provide an informed alternative to the traditional single user interface per application approach. Prior to changeable user interfaces, a single user interface was required to support all intended users. The principles of User Centred Design dictate that one should "know the user" (Nielsen, 1993) but in a single user interface, an understanding of the intended user is developed to construct average, or typical user characteristics. This might be considered design for individual similarities. This research is concerned with addressing this paradox: why is research undertaken to to understand the specific user which then practially ignores all the differentiating factors by designing for the average user, or "everyman". An alternate approach that is investigated in this work is design for individual differences as opposed to individual similarities.

1.2 Potential Solution to the Research Problem

To briefly recapitulate, this work addresses the problem of limited research into the means by which changeable user interfaces, such as user interface skins, can be used to provide a more personally relevant interaction. Furthermore, skins offer an alternative approach to a single user interface which is required to accommodate a majority of potential users.

The potential solution presented in this thesis is to investigate ways in which user interface skins can be personalised to provide a more personally relevant interaction. User interface skins offer the ability for almost any user to change their own user interface and therefore provide a means for personalisation. This is because in theory

there is the possibility for every single computer-using individual in the world to own and use their own personal user interface skin.

As part of a potential solution, this research proposes to use individual differences rather than individual similarities to help understand how changeable user interfaces can best support different user types to provide a more personalised interaction. In the field of psychology, a well researched individual difference is that of personality and it is personalisation of aesthetic elements of the user interface by personality trait that this investigation will pursue.

There is some evidence to support the position that personalisation is preferable to non-personalised, or "generic" user interfaces. Fine and Brinkman (2005) found that whilst the average highest rating was the standard Microsoft Windows skin of Windows Media Player, 36% of users rated a *Simpsons* themed skin as their most preferred skin in contrast with their rating on the Microsoft skin at 25%. This suggests that personalisation using skins provides a preferable user experience for a segment of the larger population

1.3 Vision

This section describes the high level vision and motivations for this research. It is comprised of two subsections. The first describes the vision for the context by describing current distributed computing projects that utilise high participant numbers and relatively small individual contributions to achieve large and complex computing tasks. The second section provides the vision for a design feedback loop methodology for continuous software design and development post release that is facilitated by the collection of usage data.

1.3.1 Distributed Computing

This first section describes the vision for the research context. In trying to create a more personalised user experience using any factor, data about users is required. Typically for collecting personality data self reporting questionnaires are used. However, the development of the Internet and distributed processing projects such as UC Berkeley's BOINC (Berkeley Open Infrastructure for Network Computing[1]), The SETI Institute's SETI@HOME (Search for Extraterrestrial Intelligence[2]), Stanford's FOLDING@HOME[3] and LHC@home (Large Hadron Collider[4]) all use the distributed computing power of unused or volunteered processor cycles for a single purpose. There is a diverse range of research applications of distributed computing including looking for cures for cancer, sequencing DNA or searching for extraterrestrial life, and there are many other successful distributed computing projects that leverage users freely giving up computing resources for a greater good or common goal. The success in participation for projects such as these frequently numbers hundreds of thousands of participants (e.g. BOINC 335,073 participants as of 19/01/09, Distributed.NET[5]'s cracking of the RCA-64 encryption cipher had 327,856 participants and Folding@home has over 400,000 contributing machines as of January 2009. This demonstrates that the volunteering of resources that cost the user little or nothing can be freely given if sufficient motivation is provided. It also successfully demonstrates the principle of "giving to get", with large amounts of users all contributing small amounts of resource that in total gives back knowledge and understanding to all.

It is within this context of distributed computing, and users freely giving zero-cost resources that this research sits. As was stated earlier, user data is required in order to personalise, and so if users can be persuaded that giving up a zero-cost resource, such as non-personally identifiable log data, can ultimately lead to a more personalised user experience. The main difference between the vision presented here and the distributed processing projects is that users are giving up processor cycles and not personal data. Both however do cost the user little or nothing, and can happen

[1] http://boinc.berkeley.edu/
[2] http://setiathome.ssl.berkeley.edu/
[3] http://folding.stanford.edu/
[4] http://lhcathome.cern.ch/
[5] http://distributed.net/

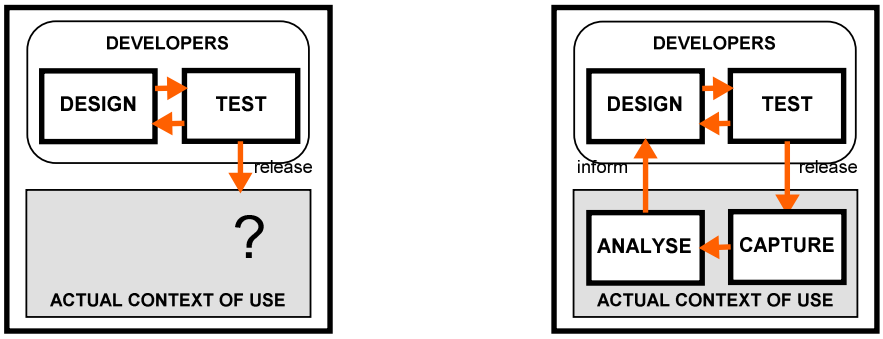
unobtrusively with little apparent effect on computing resources. It is one of the challenges facing this work to convince large numbers of potential participants that giving up user data is worthwhile, zero-cost and anonymous, in order to achieve high participation and to generate enough data to test the experimental hypotheses.

1.3.2 Design Feedback Loops

Having described the high level vision for the context, this section describes the more applied, lower level vision for a design feedback methodology that utilises ongoing usage data collection from users to redesign and redevelop software applications post release.

In practical terms, collecting data unobtrusively from users provides the opportunity for a design feedback loop. Traditionally once an interface has been designed and tested it is released for use. Feedback about users and their usage of the application is not typically logged by the application and so feedback is generated from sources such as user communities (e.g. forums, user groups), calls to a customer service department or supervisor, or even declining usage (e.g. absenteeism, increased customer churn). If interaction data was logged and sent to the developers, more understanding about users could be obtained with the option of segmenting the user population and designing skins specifically for particular user segments post release. This loop of an application generating user data, analysis of user data by developers, developing skins in light of analysed user data and then re-release of the skins, which again generates user data, provides an ability to continuously understand and improve usability and user experience post-release. Specifically for this research the improvement in the user experience suggested comes from a more personalised interaction derived from more personally relevant or preferable user interface skins. Below in Figure 1.2 are two representations of software release. In (a) the internal processes of interface design and test are represented, with revision cycles occurring between the two. However, once the software is released externally into the actual context of use, there is little direct return of usage information from the application to the developers and so the revision cycle slows and relies upon offline information. In contrast, (b) represents the vision for this research. By logging usage information

relating to the users and the user interface, the revision cycle can continue post release to provide a more optimal user experience.



(a) Traditional interface development (b) Vision for continuous development

Figure 1.2: Comparing traditional interface development and release (a) with the vision for continuous interface development and release based on captured data (b)

Having looked at the vision for this work, the next section presents the research questions.

1.4 Research Questions

This thesis presents a method for personalising interaction using user interface skins, using personality as a design segment. The method involves two main research questions. The first is: can personality be measured from interactive behaviour? This is an important question because traditional psychometric instruments tend to be time-consuming and impractical for administration. The second research question is: which user interface design elements can be linked to personality? Three types of user interface element are investigated: colour, shape and meaning. The first two are thought to be operationalised by psychophysiological arousal. Meaning in user interface skins is investigated through a testing of the Similarity Attraction Hypothesis, to determine if people choose interface skins that exhibit the same high levels of certain personality traits that they score highly in. Further detail regarding the research questions can be found in Chapter 2.

1.5 Research Approach

This research proposes an empirical approach, quantifying both personality trait and interactive behaviours for statistical analysis. Personality trait data will be obtained from the IPIP-NEO personality inventory (Buchanan et al., 1999) which is a widely accepted, robust psychometric. This instrument is based upon the Big Five model of Personality (Costa and McCrea, 1985), a dispositional approach to the study of Personality.

Internet-based research using psychometric instruments has been established as a valid methodology (e.g. Smith and Leigh (1997). Buchanan (1998), Buchanan and Smith (1999), Pasveer and Ellard (1998), Buchanan et al. (2005)). Using an externalised methodology facilitated by the Internet, a client server experimental platform was developed to investigate four experimental hypotheses. The platform features as the experimental application an Internet web radio that communicates with a central server. Participants downloaded the WebRadio application in their own context of use. The application features the ability to listen to Internet radio stations, search for stations to listen to, control volume, change user interface skins and administer questionnaires. Furthermore, the radio features a logging mechanism that enables mouse clicks relating to the web radio to be captured and recorded to a local log file. Log files are transmitted from the radio client to the central server and stored in a database, where all log files are collected for analysis. The data analysis involved investigating relationships between personality trait and both interactive behaviour and user interface skin preference using parametric statistical tests. Further detail regarding the experimental platform can be found in Chapter 3.

1.6 Thesis Overview

This section provides an overview of the thesis, providing the location and description of the positions taken in this work to investigate the personalisation of interaction using Personality-based user interface skins, starting with Chapter 2. A visual guide to this thesis is provided in Figure 1:3 overleaf.

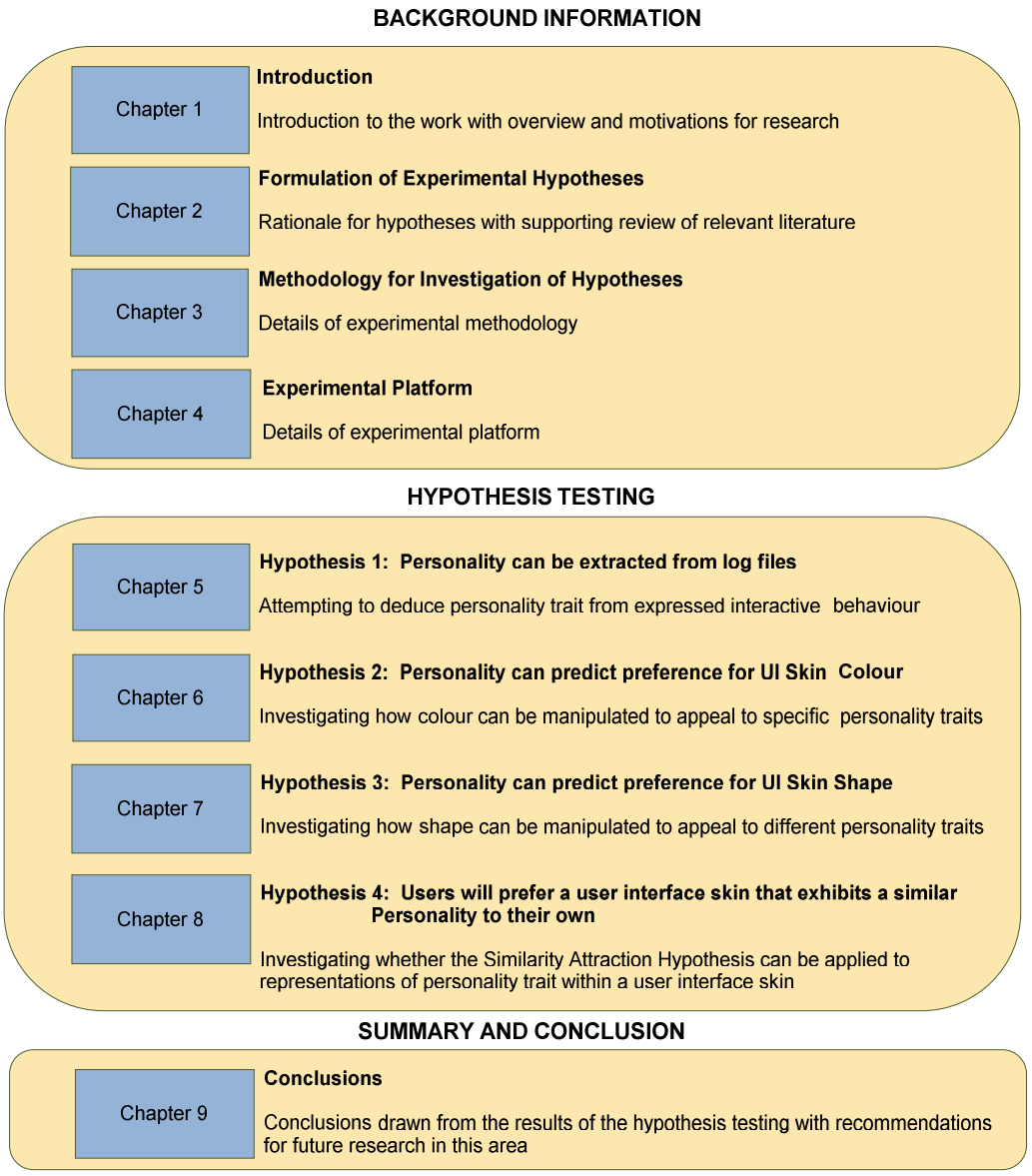


Figure 1:3 - Thesis Overview by Chapter

In investigating personalisation of the user interface, Chapter 2 serves to state and argue the positions taken in order to formulate the experimental hypotheses. Specifically, it provides the positions that establish that the user interface can be changed using the appropriate software architecture and technologies to facilitate user interface skins. The position is then taken that segmentation of large user populations by a particular defining characteristic (e.g. age, gender, personality type) can be used as a reduction strategy enabling designers to produce skins designed for a targeted

subset. The particular segment selected by this research is personality, with the position taken that personality is an appropriate individual difference to segment a large population by. Furthermore, it is argued that Personality can be linked to user interface design elements (colour, shape and meaning) so that user interface skins based on personality can be produced, thereby providing a more personally relevant interaction than the one offered by single user interface approaches. These positions lead to the stating of the four experimental hypotheses at the end of Chapter 2. The four experimental hypotheses are listed below in Figure 1::

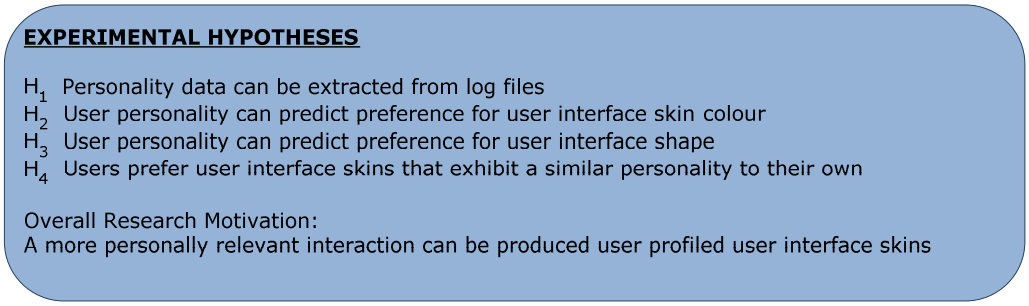


Figure 1:4 - Summary of Experimental Hypotheses

The thesis then continues by detailing experimental methodology used to investigate the experimental hypotheses in Chapter 3. Positions are argued in favour of externalised research with participants remote to the experimenter as compared with internal, experimenter-present approaches. The experimental platform and the technologies used are described in detail.

The next four chapters, Chapters 4 to 7 inclusive, are hypothesis-specific chapters. Each hypothesis is investigated in a separate chapter containing method, results, conclusions and discussion.

The final chapter concluding this thesis takes positions based upon the results of the hypothesis testing. The results indicated support for the first experimental hypothesis, that personality can be determined from log files. The second hypothesis, that colour can be linked to personality, was supported by evidence indicating Personality is a factor in preference for user interface colour. The third hypothesis, that shape can be linked to Personality, was not supported. The fourth experimental hypothesis,

investigating the Similarity Attraction Hypothesis, was supported with evidence from the extroversion trait. However, for neuroticism strong evidence for a Dissimilarity Attraction Effect was found. The overall experimental hypothesis that a more personalised user interaction can be produced using profiled user interface skins was supported through the evidence found in the experimental hypotheses, with the main exception of shape. Personality was found to be an HCI relevant design factor. The implications from these findings are discussed and reformulated experimental hypotheses are presented for future work.

To summarise, this chapter has presented the motivations and research questions that have directed this research. The next chapter formulates the experimental hypotheses through a critical evaluation of the related literature, taking positions and then providing either logical argument or literature references to support them.

Chapter 2 - RESEARCH AREA AND HYPOTHESES

2.1 Introduction

There are two main purposes of this chapter: the first is to define the research area addressed by this work, and the second is to formulate the experimental hypotheses through an evaluation and critical analysis of existing research as it relates to the personalisation of interaction.

In the opening chapter of this thesis the motivations for this work were discussed, the primary motivation being to provide a more personally relevant interaction. To recap, a more personally relevant interaction for the purposes of this research is defined as an interaction which is designed to be personally appealing to different types of user. By providing an interaction with such appeal it provides a more personally relevant interaction than a single user interface designed for the average intended user. This chapter starts out by discussing the technical and architectural means available that enable customisation of the user interface and then investigates how such interface changes can be associated with subsets of a larger population. The chapter then continues by arguing that user personality can be used to define these subsets. This allows user interfaces to be produced that appeal to certain personality traits – thus producing a more personally relevant interaction designed for types of user, as opposed to interaction designed for all users represented by the average user. The next section of the chapter looks at visual aesthetic factors (colour, edge complexity and perceived personality trait) that can be used to link user interface design to personality trait. This is done by operationalising individual factors such as physiological arousal and meaning using visual aesthetic factors such as colour and meaning in the design of user interface skins for segments based upon personality trait. The chapter then concludes with the experimental hypotheses for this thesis.

The research questions arising from and addressed by this research are summarised overleaf in Figure 2-1. This is to provide an indication of the issues that have guided this research, which provide the context for the Section Overview that now follows. These questions and their location in this chapter can be found in Figure 2-2:

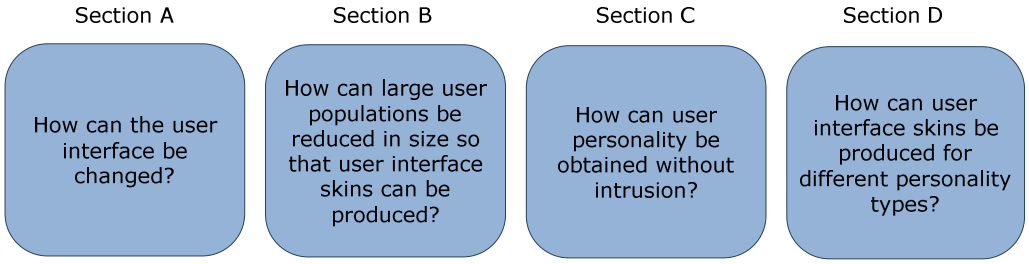


Figure 2-1 - Research Questions

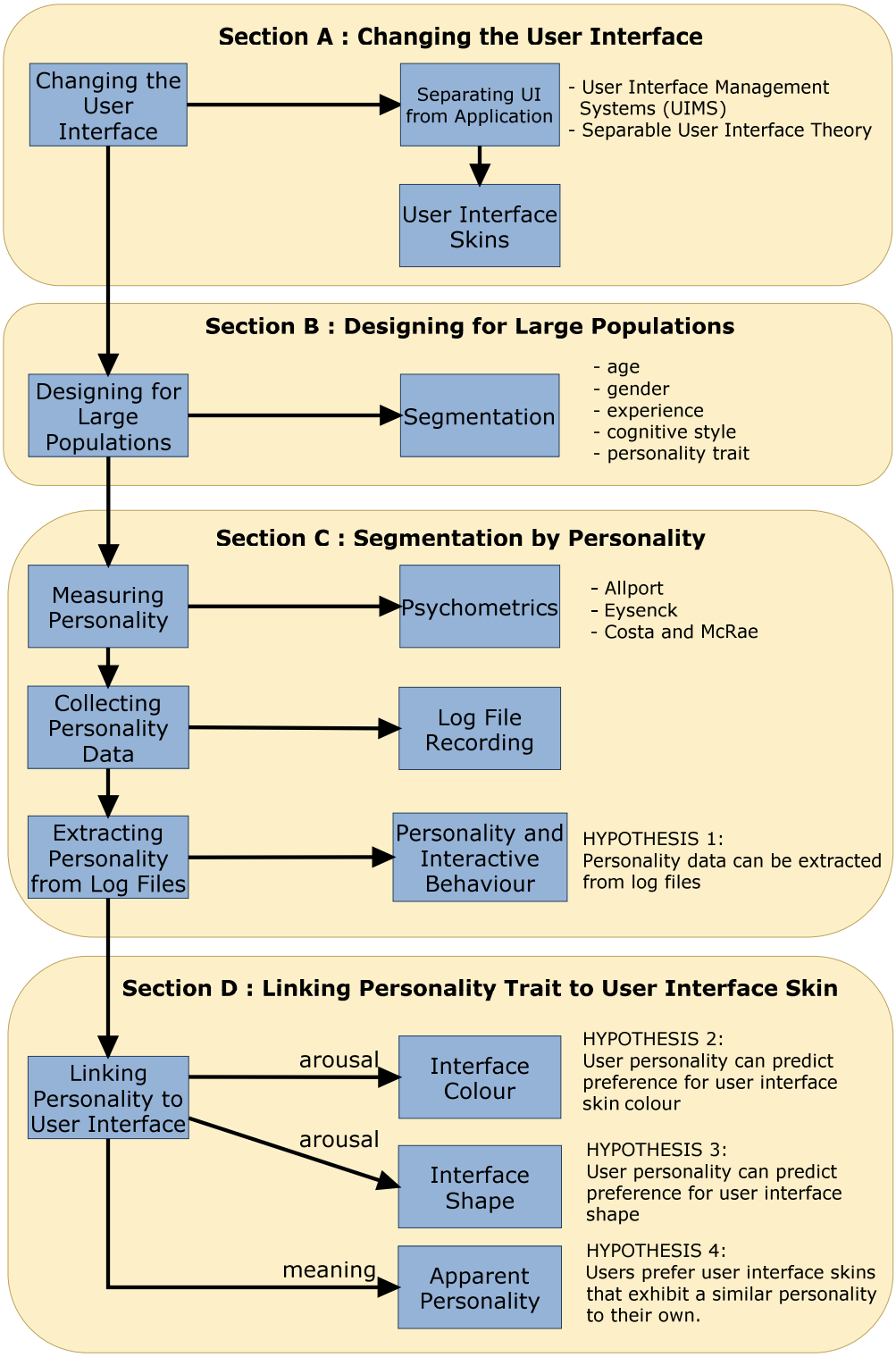


Figure 2-2 - Chapter 2 overview

2.2 Section Overview

2.2.1 Section A: Changing the User Interface

The opening section of this chapter argues that the user interface can be separated from the rest of the application logic. This is important because it allows for the user interface to be changed easily without requiring reprogramming. If one is to try to make an object – real or virtual - more personal by customising it, for example by changing the colour, one of the first questions to be asked is "how can it be achieved?" In the case of a virtual object, such as the colour of a media player, there are different means by which to change the user interface to make it more personal. Separable user interface architecture provides an appropriate means for customisation. This section argues for the separable user interface by providing supporting evidence of the architectural means by which the user interface can be changed, and how this facilitates customisation of the user interface.

2.2.2 Section B: Designing for large user populations using segmentation

Having covered the means available for customisation, the next section addresses the next logical question in the process - how can interaction be customised for large user populations? This section argues that large populations can be segmented into smaller subsets defined by HCI-relevant factors (e.g. age, gender, cognitive style). As the opening chapter of this thesis notes, designing for the individual can be an impractical and relatively resource-heavy task, and designing for the average user presents an average interaction of limited personal appeal. This section addresses how large populations can be segmented into smaller subsets, defined by a common factor (e.g. age, gender, cognitive style).

2.2.3 Section C: Segmentation by personality trait

As seen in Section B, there are a large number of potential factors that can be used to segment large user populations. This section focuses on one factor, user personality type, and explains why personality might be a useful factor by which to segment large user populations. Logically, in order to provide a more personally relevant interaction, the personality type of the user may be considered as a factor governing personal appeal. This premise is supported by a number of recent studies which are

summarised in this section, providing support for the use of personality data in the process of segmenting large user populations.

The section continues by addressing how personality data can be collected and argues in favour of unobtrusive data collection using log file recording. A most frequently used instrument to measure personality is that of the self reporting psychometric questionnaire (Cattell, 1956, 1965; Eysenck, 1961; Goldberg, 1999; Costa and McCrae, 1999). However, presenting users with a Personality inventory to complete prior to interacting with a piece of software can be impractical and unrealistic because of the interruption or delay in the primary task of interacting with the software itself. This presents a potential problem for research centred on using personality as a customisation factor. In contrast to the explicit data collection method employed by self-reporting questionnaires, log file recording provides an alternative means by which to collect personality data without interrupting the primary task. This section takes the position that it may be possible to extract personality data from behavioural data captured by log files, negating the need to involve the user directly in the collection of personality data. This provides an opportunity to determine user personality from recorded interactive behaviour and leads to the first hypothesis, (H_1, Figure 2:3), that personality can be established from behaviour captured in recorded log files.

2.2.4 Section D: Designing user interface skins for personality types

Having established a means to collect Personality data in an unobtrusive manner, this section investigates potential visual aesthetic factors available for manipulation in user interface skin design so that profiled user interface skins can be designed. Three factors in total are investigated; colour, shape and perceived skin personality.

The two main aesthetic factors investigated are colour and shape. A profiled user interface skin (ProSkin) is a user interface skin designed for a particular profile of user, in this case the profile defined by personality type. This section argues that psychological customisation (using ProSkins) can be used to create a more personally relevant interaction, because the user interface skin is of relevance to the personality type of the user. The section explores the use of colour and shape in design and

examines any potential interaction effects between skin colour (and shape) and personality trait and the effect on interactive behaviour.

Whilst it is not within the scope of this research to determine exactly what mechanisms are responsible for behavioural change, the theory of the underlying mechanisms are summarised in this section (e.g. physiological arousal, semantic representation, cultural variance). Although there is relatively little research into the effect of visual aesthetics on different personality types, there are studies that suggest relationships between user interface properties and physiological arousal (Berlyne, 1994; Martindale et al., 1994). Other studies have found links between the personality trait of extraversion and physiological arousal, so it is argued that there is a logical relationship between personality trait and user interface properties. The two aesthetic factors affecting arousal are researched in this thesis, user interface colour and user interface shape, and these are represented by Hypotheses 2 and 3 (H_2 and H_3, see Figure 2-3).

An additional third factor, perceived skin personality, is also investigated. The Similarity Attraction Hypothesis (Byrne and Nelson, 1965) suggests that people prefer those who are similar to themselves, so this section argues that the perceived personality of the user interface skin can be a predictor of user skin preference. The fourth experimental hypothesis (H_4) investigates whether the Similarity Attraction Hypothesis holds true when applied to HCI by representing personality traits using images that express a specific personality trait (e.g. extraversion). Whereas the first two interface factors (colour and shape) operationalise user arousal, this third factor operationalises meaning within the skin – in this instance the perceived personality type represented in the skin. This is to investigate whether an additional non-aesthetic factor such as meaning has effects beyond simple aesthetic factors (colour and shape).

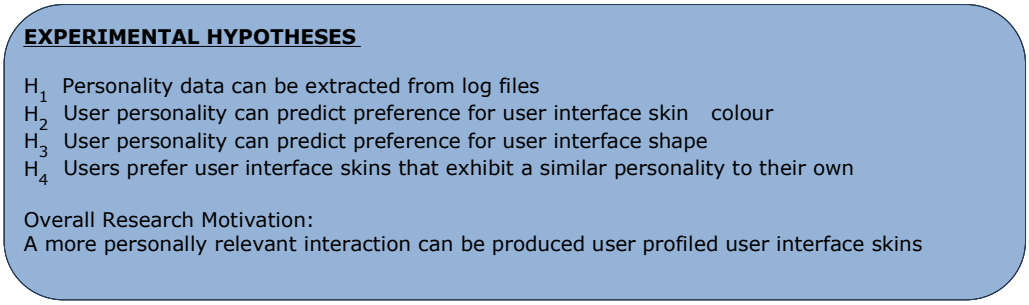


Figure 2-3 - Experimental hypotheses with overall research motivation

2.3 SECTION A: Changing the user interface

This section argues that architectures exist that enable the user themselves to change their own user interface quickly and easily at runtime, without the need for reprogramming of any application logic. Evidence to support this position is provided in this section by referencing software architectures that enable a separable user interface (e.g. Model-View-Controller (MVC), Seeheim). It is important that the user interface can be changed easily, because without this factor a customised interaction for large user populations becomes impractical. The section begins by briefly describing the unchanging interface, referred to as a static interface for the purpose of this thesis. An understanding of the software architectures prior to and post static user interfaces is important because it provides the basis and context for changeable user interfaces, the central theme for this research. Figure 2-4 below describes the structure for this section:

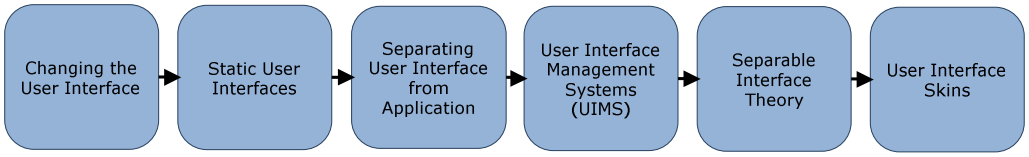


Figure 2-4 - Section A structure

2.3.1 Static user interfaces

Historically the user interface has been a static entity and typically software applications have featured a single user interface to accommodate all users (Myers, 1998). This has been because the user interface was coded alongside application logic and as such "hard coded" within the system and therefore only modifiable by developers with access to the source code. The end result was that there was no architectural distinction between the user interface and the application code, and this meant that any changes to the user interface became non-trivial tasks. Such static user interface architectures have limited the amount to which user interfaces could be varied because of the effort required to change them. In trying to provide a more personally relevant interaction using customised user interface skins, low effort in changing the user interface is important because it facilitates widespread application. The next section addresses software architectures that facilitate low effort customisation.

2.3.2 Separating User Interface from Application

There are a number of different software architectures that define the user interface separately from the application (Seeheim, MVC, User Interface Management Systems (UIMS), and Separable UI Theory). The Seeheim model (Green, 1983) introduced into software engineering the concept of separating content from presentation. It provided a model of the user interface as distinct from the rest of the application because it represented the required components for automatic user interface generation.

The MVC paradigm for software architecture (Reenskaug, 1979) explicitly separates the data model, application logic and user interface of an application. This allows changes to be made to one component without requiring significant change to the other components. This is important because the user interface is defined architecturally as separate from the rest of the application logic (see Figure 2-5). The component marked 'View' represents the user interface.

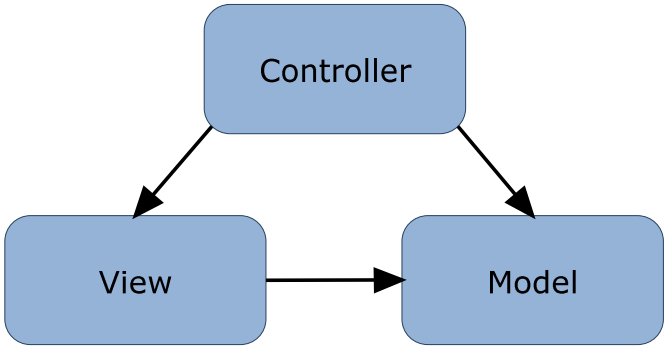


Figure 2-5 - The Model-View-Controller

Software architectures such as these that define the separation of user interface from application logic have become known collectively as User Interface Management Systems (UIMS).

2.3.3 User Interface Management Systems (UIMS)

UIMS have been used as an architecture that enables changing the user interface on multiple different applications, enabling a single user interface to be generated for multiple applications (see Figure 3). However there is little agreement in the literature as to the definition of UIMS. Some regard it to represent the distinction between application and interface and therefore refers to the user interface itself (Evers, 1999; Hurley and Sibert, 1999; Rosenberg et al., 1988). Others define UIMS as analogous to database management systems (DBMS) in that both UIMS and DBMS provide levels of abstraction to automate time-consuming tasks, for example the automatic generation of the user interface (Foley, 1991; Myers et al., 2000). The importance of UIMS is of relevance to this research because it provides a concept of the interface as distinct from the rest of the application so as to automatically generate a user interface. Separable User Interface theory (Edmonds, 1992) provides another means by which to automatically generate user interfaces, although from a very different architectural viewpoint to UIMS. It is the architectural differences between UIMS and Separable User Interface Theory that facilitates user changeable dynamic user interfaces.

2.3.4 Separable User Interface Theory

Edmonds (1992) developed the concept of the UIMS into a theoretical framework for creating user interfaces that emphasises the independence of the user interface from the functional aspects of the application that the interface serves. This framework was called Separable Interface Theory and it was intended that it could lead to the automatic construction of user interfaces (Green, 1985). Separable Interface Theory differs from UIMS because it facilitates multiple user interfaces for a single application, unlike UIMS which facilitates a single user interface for multiple applications. The architectural differences between UIMS and Separable User Interface Theory are illustrated below in Figure 2-6. Recent software engineering developments that facilitate separable interface theory include Extensible Markup Language (XML) and Cascading Style Sheets (CSS), because both provide a means by which to define and separate data according to presentation and content functions. The actual technologies involved in developing separable user interfaces is not within the scope of this work, however those that are directly related to the development of the experimental environment for this research is discussed in detail in Chapter 3.

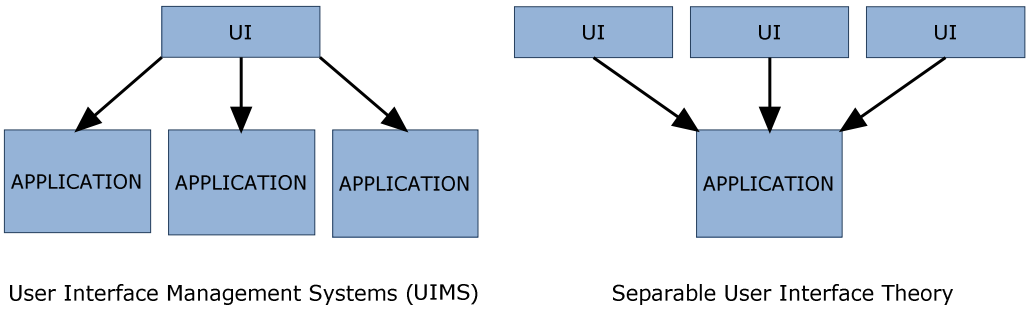


Figure 2-6 - Architectural differences between UIMS and Separable User Interface Theory

2.3.5 User Interface Skins

One application of Separable Interface Theory are User Interface Skins, colloquially abbreviated to "skins", the name implying the user interface as a layer of the application apparent to the user. The definition of the term "skinning" varies but most include the concept of changing the user interface skin to something preferable to the

user. For the purposes of this thesis, skins are defined as representing the aesthetic elements of the user interface and can include visual, aural and haptic elements. Skinning (or "reskinning") an application facilitates the provision of multiple user interfaces for a single application. It is because of this ability that they are widely used in the customisation of the user interface.

In order to illustrate how user interface skins can change the appearance of the user interface, Figure 2-7 below shows different coloured skins on the same application (in this case a media player called WinAmp). In this first example the only factor that changes between the skins is colour.

Figure 2-7 - Single application, multiple user interface skins (colour only variation)

This exemplifies one of the simplest ways to customize the user interface by changing the colour palette. None of the interface components change, for example button sizes or position, and only the colours used vary. For the purposes of this thesis, skins that feature only variation in colour are often referred to as "colour themes" or "styles" but for the purposes of this research are considered to be user interface skins because of the hypothesized relationship between user interface skin colour and interactive behaviours exhibited by different personality traits.

Another means by which to customize the user interface is to change the interface components, for example the play button, the song display window or the shape of the application itself. Four examples of this kind of customization are illustrated below in Figure 2-8. These images are all user interface skins for the same application as in Figure 2-7, yet their appearances are radically different from one another. The

application itself still has the same functionalities available but the user interface skins provide both different appearance and different interaction. For example in Figure 2-7 all the interface components are positioned the same and only the colour palette changes. In contrast, the example skins shown in Figure 2-8 illustrate how changing the colour, shape and position of interface components can present the user with a different interaction with the same application. This is because the user interface components in each skin have been changed to provide a different aesthetic appeal. Note that there are no standard positions for any of the interface features in Figure 2-8 (e.g. the play button varies in position in each skin).

Figure 2-8 - Single application, multiple user interface skins (full variation)

These two examples show how user interface skins can be used to customize an application. There are other modalities that can be catered to, such as auditory or haptic (for example, by changing system sounds) but this research focuses on the visual aesthetics factors required to produce a more personally relevant interaction.

The next step in developing a more personally relevant interaction using user interface skins is to understand the nature of the intended users, so as to produce user interface skins that are of relevance to them. The next section addresses how large user populations can be reduced to smaller subsets by segmenting users according to common factors.

2.4 SECTION B: Designing for large populations using segmentation

This section argues that large populations can be segmented into smaller subsets defined by common factors (e.g. age, gender, cognitive style). There are a number of different ways that a segment of users can be described. This research focuses on factors that have relevance to the human-computer interaction – referred to in this thesis as HCI-relevant factors. An HCI-relevant factor is defined here as any individual difference that has a direct effect on the human-computer interaction, as opposed to non HCI relevant factors which do not have clear relevance to the interaction (e.g. hair colour, salary). The reason for the exclusion of such factors from this research is because it is difficult to accommodate such physical or non-HCI relevant factors within a user interface skin in a software application. In trying to determine how best to segment a large user population it seems sensible to choose factors that can be accommodated in the design of user interface skins.

Before looking into the ways in which user interface skins can be best applied to different user segments it is first necessary to understand something of the intended user population. As stated in Chapter 1, it is often impractical to design for the individual, so large user populations can be redefined in terms of subsets of user characteristics, in order that user interface skins can be designed for common characteristics. For example, one of the most basic ways to segment a large population is by gender, which reduces a large population to two subsets; male and female. This could mean that there are user interface skins that provide a more personally relevant interaction for males, or females. This distinction is a simple one, but even at a simple level it provides an opportunity to provide an improved interaction for large user populations using only two skins (male and female). Figure 2-9 overleaf provides an overview of the structure for Section B:

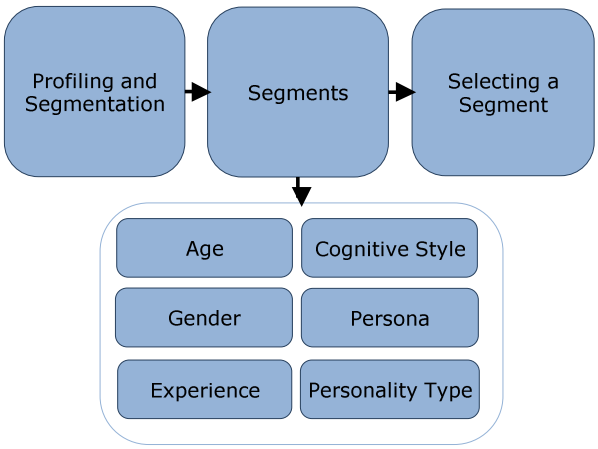


Figure 2-9 - Section B structure

Before discussing the means by which user populations can be segmented, it is first necessary to define the concept of segmentation, and to differentiate it from user profiling.

2.4.1 Profiling and segmentation

It is important to differentiate between the concepts of profiling and segmentation, because this research uses segmentation as opposed to user profiling as a means to develop user interface skins. User profiling is a methodology by which information is gathered about an individual user to form a profile of that user, often with the intent of providing personalized content (Adomavicius and Tuzhilin, 2001; Pazzani, 1999). For example, a user profile may include information such as age, gender, location, interests and historical behaviour. The profile is created exclusively for the individual. Segmentation, unlike profiling, is not applied at the individual level (Filgeuiras and Aquino Jr, 2005). It is used to subdivide a group according to a common or shared characteristic so as to produce segments for which user interface skins can be designed.

For the purposes of this research, a modified definition of profiling is used. Profiled user interface skins (abbreviated to ProSkins) are skins that have been created according to a profile, and that profile has been determined through segmentation. Therefore the profile refers to a segment profile, rather than an individual profile. An example of this is an age-related ProSkin to assist elderly users. The profile could be

defined by any users of either gender with an age greater than 65 years of age, and this then forms the segment. Potential user interface features that would better support this segment might include increased font sizes and a higher contrast colour palette (Czaja and Lee, 2003).

There are a number of possible factors that can be used in the segmentation of a large user population. These include:

2.4.1.1 *Age*

Initially age might seem like an obvious way to segment, but age represents a multitude of factors, including experience (e.g., development of mental models), cognitive ability (e.g., processing time), physical ability (e.g., eyesight). Age-related changes include loss in near vision (Czaja and Sharit, 1998), reduced field of vision (Hawthorn, 2000), decreased colour and contrast sensitivity (Helve and Krause, 1972), and hearing reduction or loss (Kline and Scialfa, 1996). Measuring age represents a large set of factors and makes it difficult to isolate the dependent variable causing the experimental effect, for example psychological, cognitive, physiological, cultural and social. However, the range of factors included described by age provide a practical means by which to segment large user populations. Research into the effect of age on interaction by Docampo Rama (2001) found that a generational effect could be observed and predicted interactive success. Two main generations were found; an electro-mechanical generation (born prior to 1960) and a software generation (born after 1960). These generational differences show that different ages interact differently and therefore provide a logical argument supporting interface design by user segments, such as age, or generation.

2.4.1.2 *Gender*

Gender is a social construct (Haralmbos, Holborn and Heald, 2006) that may also provide a useful means to segment large populations. There is some emerging research relating to gender differences in HCI. Brinkman and Fine (2005) found gender differences relating to different user interface skins in two different studies. In the first study their results suggest that females have a tendency to rate skins featuring 'cute' characters significantly higher than males, and conversely males rate military themed skins significantly higher than females. In addition to this their results also

suggest gender differences in colour preference for user interfaces, with females rating blue, purple and white skins significantly higher than males. Such gender differences in preference for user interface skin were also found in a second study, where females rated skins in a category of "small furry creatures" significantly higher than males. Males rated skins from categories "small skins", "medium skins", "square shaped skins", "predator skins" and "scary skins" significantly higher than females. These two studies both suggest that gender differences provide a means by which to segment a large population.

In addition to Brinkman and Fine (2005), Beckwith (2005) found gender differences in the error rates of male and female users in the context of debugging spreadsheets. Czerwinski, Tan and Robertson (2001, 2002, 2003) found gender-specific differences in the users field of vision, with females benefiting from optical flow cues that wider screen displays provide, without any apparent impact upon male usage of the same display. Although limited, this research may indicate that gender differences can be accommodated within interaction design by providing gender–differentiated segments for profiling.

2.4.1.3 *Experience*

The amount of time that the user has previously spent using an application is another means by which to segment (Mason, 1986; Docampo Rama, 2001). A typical distinction made is that of Expert and Novice user. For example, a menu system can be run in 'verbose' mode that provides context-related help for the inexperienced user, at the cost of more text on the screen. For the more experienced user who no longer needs contextual help, a 'brief' mode reduces the menu to the main functions, reducing task completion speed. User experience can be observed and measured both qualitatively (user questionnaires) and quantitatively (task completion time). Work by Docampo-Rama (2001) found that age and experience factors were predictors of interactive performance. Her work suggests that experience with a software application leads to a well trained visuo-spatial working memory, and that this improved cognitive ability is responsible for improved interaction resulting in experienced users. Individual differences in experience may define a user segment from which a profile can be created and supported in the user interface, with

functionality limited to core features for novices and advanced features enabled and available for experienced users.

2.4.1.4 *Cognitive Style*

Cognitive style is defined as "an individual's habitual way of perceiving, remembering, thinking and problem solving" (Riding and Rayner, 1998 p.15). They are high-level heuristics that organize and control behaviour across a wide variety of situations (Dufresne and Turcotte, 1987). Dufresne and Turcotte state that cognitive styles are important in determining the most effective interaction style, and that Field dependence is the cognitive style associated with the greatest amount of research conducted. Field dependence is derived from form the work of Witkin et al., and relates to the ability to abstract items in a specific context as well as affecting the way a person structures and processes information (Witkin et al, 1962). Dufresne and Turcotte observed different navigation strategies in users navigating spreadsheet software as a function of field independence. They found that Field independent users appeared to browse the interface first to construct a mental model of the content and then to return for a more focused interaction, and that Field dependent users did not explore the system to the same extent as Field independent users had. These findings replicate the findings of Coventry (1989) and provide evidence that cognitive style may be a valid means to segment large user populations, as it appears that human computer interaction varies as a function of cognitive style.

Benyon and Murray (1993) designed systems which adapted to individual differences in personality and cognitive style. This was performed by identifying individual characteristics, validating them and then designing appropriate interaction methods that accommodated these individual characteristics. They measured spatial ability, verbal ability, field dependence, logical/intuitive thought and short term memory ability in their experimental participants and then asked users to perform interactive tasks in each of five different user interfaces. They found that the largest differences in performance in the command line interface were a function of spatial ability, and that menu style interfaces are more suitable to low spatial ability users than command line interfaces. Their work shows suggests that cognitive style can produce interaction differences that can be accommodated by user interface skins to provide a more personally relevant interaction.

Chen and Macredie (2002) conducted a comprehensive review of the research into cognitive style and the effect on learning using hypermedia navigation. Having analysed the body of existing research they constructed a model for learning based upon the Field dependence/independence cognitive style. The model suggests that different cognitive styles produce different navigational strategies and that the user interface should support differing cognitive styles in the user. This suggests that cognitive style might be an appropriate means by which to segment.

2.4.1.5 *Personas and Stereotypes*

One means by which to describe users is to use personas or stereotypes (Pruitt and Grudin, 2003; Cooper, 1999). These are representations of types of intended user that can be based upon fact or fiction (Cooper and Reiman, 2003). An issue that commonly affects commercial software development is that once an application has been released for general usage there is often very little data feedback to the application or interface designers to inform them as to how the application was actually used. This can mean that designing using stereotypes is frequently based upon focus groups that attempt to represent the intended user population. However, the use of personas does not provide an effective means to segment large user populations. This is because personas are frequently derived from the average or the typical intended user and does not consider the individual. This research argues that user interface design should be away from the individual, but also away from average, and towards interaction for types of user (profiles derived from segments). The use of personas and stereotypes by definition is inherently based upon average. The use of segmentation in this research is so that identifiable individual differences common to a subset of users can be used to define a segment, for whom profiled user interface skins can be produced.

2.4.1.6 *Personality*

Within psychology the area of personality has been extensively researched and provides another means to segment large user populations. There are a number of different approaches to the study of personality, including the Psychoanalytical (e.g. Freud), Behaviourist (e.g. Skinner) and Humanistic (e.g. Maslow). This research focuses on the dispositional approach, which is based on the concept of stable user

characteristics or traits that are expressed in behaviour for different given situations (Pennington, 2003). This approach sees the individual as characterised by a unique configuration of traits. Personality traits are considered to be "the basic qualities of the person" (Mischel, 1993 p.38) and are a temporally stable, cross-situational individual difference. This stability over time is important because it allows personality factors to be used to segment large user populations reliably over time. Using factor analysis statistical techniques, Eysenck (1947) was one of the first psychologists to measure personality factors. Initially two personality dimensions were identified, extraversion and neuroticism, but psychoticism was later added as a third (Eysenck and Eysenck, 1976). A well researched variant of Eysenck's work in dispositional research is the "Big Five" model of personality (Costa and McCrae, 1985; Goldberg, 1990) In addition to Eysenck's three factors another two dimensions are added making up the "Big Five"; openness to new experience, conscientiousness, extraversion, agreeableness and neuroticism (see Figure 2-10 overleaf). There are five main factors each with six sub factors making a total of thirty dimensions measured. They can be measured using psychometric tests and the next section specifically addresses the collection of personality data. This is because in this research personality factors are used to segment users, so as to provide user interface skins appropriate to the factors common to that segment, i.e. personality factors such as extroversion.

OPENNESS TO NEW EXPERIENCE
Assesses proactive seeking and appreciation of experience for its own sake; toleration for and exploration of the unfamiliar.

Sub Factors:
Imagination, Artistic Interests, Emotionality, Adventurousness, Intellect, Liberalism.

High Scorer: curious, broad interests, creative, original, imaginative, untraditional
Low Scorer: conventional, down-to-earth, narrow interests, unartistic, unanalytical

CONSCIENTIOUSNESS
Assesses the individual's degree of organisation, persistence, and motivation in goal-directed behaviour. Contrasts dependable, fastidious people with those who are lackadaisical and sloppy.

Sub Factors:
Self Efficacy, Orderliness, Dutifulness, Achievement Striving, Self Discipline, Cautiousness

High Scorer: organized, reliable, hard-working, self-disciplined, punctual, scrupulous, neat
Low Scorer: aimless, unreliable, lazy, careless, lax, negligent, weak-willed, hedonistic

EXTRAVERSION
Assesses quantity and intensity of interpersonal interaction;activity level; need for stimulation; and capacity for joy.

Sub Factors:
Friendliness, Gregariousness, Assertiveness, Activity Level, Excitement Seeking, Cheerfulness

High Scorer: sociable, active, talkative, person-oriented, optimistic, fun-loving, affectionate
Low Scorer: reserved, sober, unexuberant, aloof, task-orientated, retiring, quiet.

AGREEABLENESS
Assesses the quality of one's interpersonal orientation along a continuum from compassion to antagonism in thoughts, feelings and action.

Sub Factors:
Trust, Morality, Altruism, Cooperation, Modesty, Sympathy

High Scorer: empathetic, considerate, friendly, generous, helpful
Low Scorer: unfriendly, argumentative, unhelpful, inconsiderate, suspicious, manipulative

NEUROTICISM
Assesses adjustment versus emotional instability. Identifies individuals prone to psychological distress, unrealistic ideas, excessive cravings or urges, and maladaptive coping response.

Sub Factors:
Anxiety, Anger, Depression, Self Consciousness, Immoderation, Vulnerability

High Scorer: worrying, nervous, emotional, insecure, inadequate, hypochondriacal
Low Scorer: calm, relaxed, unemotional, hardy, secure, self-satisfied

Figure 2-10 - The "Big Five" personality dimensions, with sub-factors (Costa and McCrae, 1985)

2.4.2 Selecting a segment

The purpose of this section has been to outline methods for segmenting large user populations. Whilst this is not an exhaustative list, it provides evidence to show that a large user population can be reduced in size by segmenting users of a common trait (e.g. age, cognitive style, personality). In trying to find more personally relevant interation, a segment needed to be chosen for use in this research. Personality was selected as the segmentation factor for this research because personality traits have been shown to be stable over time (Watson and Walker, 1996; Moon and Nass, 1996) and because there is very little research available that investigates the effect of personality trait on interaction.

Having selected a means by which to segment, the next section examines the collection of personality data and argues that personality data can be collected unobtrusively from the user.

2.5 SECTION C: Segmentation by Personality

In order to segment according to user personality type a means to measure user personality is required. This section looks at the ways in which personality data can be collected using psychometrics such as the IPIP-NEO (Goldberg, 1999) and then looks at practical issues involved with collecting personality data, arguing in favour of implicit (rather than explicit) collection of personality data. The figure below provides a brief overview of the structure to this section.

Figure 2-11 - Section C structure

2.5.1 Psychometrics

Personality is typically measured through the use of self-reporting psychometric questionnaires. Questionnaires are a method used for the elicitation, recording and collection of user information and their use in psychology, computer science and HCI is an accepted and valid means for obtaining user information. Typical uses for questionnaires include user satisfaction, attitude, opinion and psychometric assessment. The main advantage of using questionnaires is that they provide a means by which to indirectly measure psychological factors from the user's point of view.

One such psychometric test, Costa and McRae's NEO-PI-R (Costa and McRae, 1996) measures the "Big Five" personality traits, of openness, conscientiousness, extroversion, agreeableness and conscientiousness and is widely accepted because of its environmental and longitudinal validity (Furnham, 1997). However, the NEO-PI-R is a commercial psychometric and presents practical limitations for academic research. Goldberg (1999) has developed a public domain psychometric to measure the FFM called the IPIP-NEO. The IPIP-NEO correlates highly with the NEO-PI-R in measuring similar constructs (Goldberg, 1999) and as such is suitable for research purposes for measuring the five factors. This provides an experimental means to practically measure personality so that the associated personality traits can be used to segment a large user population.

There are alternative means to determine personality, such as by direct observation of user behaviour. However, Azjen (2005) suggests that observational methods are limited and makes an argument against their use on the grounds that they are frequently costly and time consuming. As shall be argued in the next section, implicit data collection (using log files) provides an opportunity to observe interactive behaviour without the need for participants to be physically present in an observation lounge, thereby helping to limit the cost and time resources required.

2.5.2 Explict and Implicit Data Collection

Having established an appropriate means to measure personality, a means to administer the test to participants is required. There are two main ways to collect

personality data from users: explicitly and implicitly. This section argues that implicit data collection is preferable to explicit data collection for the purposes of this research.

Explicit collection of personality data can be done a number of different ways, but the most common tend to be self-reporting methods (e.g. questionnaires) and observational methods (e.g. questionnaire reporting by a third party). Such methods allow the investigator to measure personality by explicitly involving the user in the process. However, because they involve the user explicitly they present practical problems for this type of research. This is because the act of collecting data to establish the user's personality type is a time consuming process and interrupts the main task at hand. For example, it would be impractical to ask users to complete a psychometric questionnaire every time they wanted to interact with an application that required a concept of user personality type (e.g. customising interaction using user interface skins).

In contrast to explicit data collection, implicit data collection does not directly involve the user and minimises user involvement by passively observing behaviour (Kelly and Teevan, 2003; Fox et al., 2005). Implicit methods (e.g. log file recording) provide a means by which to observe user behaviour without interrupting the primary task to collect data. Fox et al. (2005) found that there was an association between implicit measures of user activity and the user's explicit satisfaction ratings, suggesting that there is a tendency for users to prefer implicit, less intrusive data collection means. As such implicit data collection methods provide an opportunity to extract personality data from recorded log files because certain personality factors exhibit associated behavioural patterns. The experimental evidence to support this claim is covered in the next section. The specific methodological detail is covered in Chapter 3.

2.5.3 Extracting Personality Data from Log Files

This research provides another means for understanding personality by removing the need for self-reporting. By establishing a relationship between personality and interactive behaviours, an instrument can be constructed by observing interactive behaviours as they relate to a particular user interface skin. This may remove the need for self-reporting and addresses the limitations of explicit data collection

methods. This can be achieved because personality factors are exhibited in user behaviour, and these user behaviours are captured through log file recording. Therefore if user behaviour can be linked to personality factors, personality can be constructed from user behaviour.

There are a number of findings that support this line of thought, that users express their personality in their observable behaviours. Brebner and Cooper (1974) observed a higher level of motor activity in extroverts as compared to introverts. They found that extroverts press buttons at higher rates than introverts when it causes a change in visual stimulation. Similarly Doucet and Stelmack (1997) found positive correlations between extroversion and speed of movement. Saati, Salem and Brinkman (2005) found that extroverts tended to interact faster with the user interface than introverts, and replicates reported observations of correlations between extroversion and the speed of human movement (Doucet and Stelmack, 1997). Correlations between interactive speed and conscientiousness as well as neuroticism were observed. These findings further support the view that user interface skins can be developed that accommodate personality factors to produce a more personally relevant interaction. Geen (1984) found that extroverts chose more intense noise levels than introverts when selecting a background level to accompany a learning task. These results suggest that extroverts have a greater need for variation and intensity of stimuli than introverts. This provides an opportunity to determine personality from behaviour captured by log files because certain personality traits, such as extroversion, may influence interactive behaviour. For example the variation in stimuli sought by extroverts might be observed in interactive behaviour as, for example higher frequency in skin changes than introverts (low scoring extroverts) or preference for brighter coloured skins.

These studies provide evidence to support the position that user personality can be determined from behaviour. This is because of the potential behavioural differences between users as a function of personality trait. These personality traits and the differences in interactive behaviour provide the means for segmentation by personality type. This position is represented by the first experimental hypothesis to be investigated by this research:

HYPOTHESIS 1: Personality data can be extracted from log files.

2.6 SECTION D: Linking Personality to User Interface Skin Properties

As has been stated previously, the main aim of this research is to determine how a more personally relevant interaction can be produced using user interface skins. The previous sections have described how the user interface can be changed using skins, how large user populations can be designed for by segmentation, segmentation by personality trait, and extracting this data from recorded log files implicitly. In order to be able to produce user interface skins of relevance to specific personality traits it is now necessary to be able to link user personality to user interface skin properties. This is done by varying user interface skin properties so that they are preferable to the user than a non-profiled user interface (i.e. standard user interface for all users).

This section addresses how this can be achieved and investigates three specific user interface skin properties; colour, shape and meaning. It will be argued that these factors can be manipulated to have greater appeal to specific personality traits. These positions lead to the experimental hypotheses that specific user segments (defined by personality trait profiles) can be targeted with user interface skins featuring aesthetic factors that appeal to the personality trait profile, and this then produces a more personally relevant interaction. These hypotheses are presented and outlined at the end of this chapter. The experimental methodologies, results, analysis and discussion for each experimental hypothesis are located in the following 4 chapters. Figure 2-12 below illustrates the section structure:

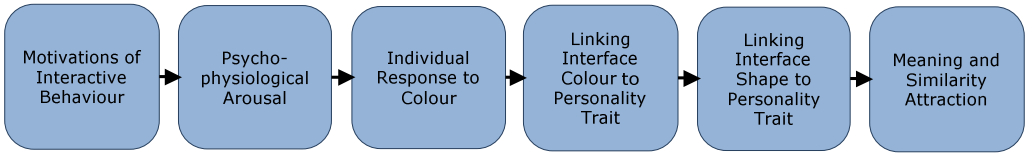


Figure 2-12 - Section D structure

2.6.1 Motivations behind Interactive Behaviours

Whilst it is beyond the scope of this research to investigate the specific motivating factors that influence interactive behaviour, it is necessary to identify them here

because they provide support to justify the user interface skin factors selected to be researched. This is because the experimental factors selected (colour, shape, meaning) are designed to operationalise these underlying motivations so that a more personally relevant interaction can be achieved. Two specific mechanisms that influence interactive behaviours are investigated in the next sections; psychophysiological arousal and meaning.

2.6.2 Psychophysiological Arousal

The literature suggests that the first two of the experimental factors, colour and shape, affect psychophysiological arousal. The observed effects of stimuli such as colour and shape on user arousal is handled separately in the next section. This section provides a brief overview of the research relating to arousal with the purpose of outlining possible reasons for individual differences in arousal responses. This helps to make the link between user personality traits and operationalised arousal factors such as colour and shape.

Psychophysiological arousal is defined for the purposes of this research as both a physiological and psychological state that arises from stimulation of the reticular activating system in the brain (Duffy, 1962; O'Gorman, 1977; Mandler, 1992). Early research by Eysenck (1947) suggested that the underlying physiological mechanism influencing extroversion is arousal level in the brain. He considered the extrovert to be less aroused than the introvert and therefore seek external stimuli to compensate for this loss. Conversely he considered the introvert over-aroused and therefore avoiding any arousal inducing external stimuli. Hebb (1955) also describes behaviour in terms of arousal and suggests that individuals are motivated to find an optimal level of arousal. If the user is underaroused then arousal is perceived as rewarding and conversely if overaroused then a decrease in arousal is perceived as rewarding. This appears to develop the ideas of Yerkes & Dodson (1908) which indicates an inverted "U" relationship between arousal and performance, which suggests that there is an optimal level of arousal required for optimal performance. Similarly, Zuckerman (1979) attributes sensation seeking behaviour to achieving arousal homeostasis. Furthermore, Berlyne (1960) observed a relationship between arousal and task complexity, with user arousal increasing as a function of task complexity.

Such studies attempt to describe the mechanisms for psychophysiological response to stimuli. It is this variation in individual response to experimental variables (e.g. colour and shape) as a function of personality trait that facilitates segmentation by personality trait and ultimately interaction design for Personality.

The next sections examine the effect of colour and shape on arousal.

2.6.3 Individual Response to Colour

This section examines the individual response to colour and argues that colour elicits an individual arousal response. In the following section this individual arousal response is then linked to personality traits (e.g. extroversion).

There is research that indicates that the perception of colour can have a subjective affect on the individual. For example there is much evidence to suggest that bright colours enhance cortical arousal in the brain (Granger, 1955; Edwin et al., 1971; Jacobs and Hustmeyer, 1974; Kreitler and Kreitler, 1972). Similarly Valdez et al. (1994) found that green-yellow, blue-green and green colours were perceived as the most arousing and purple-blue and yellow-red the least arousing.

Wolfson and Case (2000) investigated the effect of background colour and system sounds on gaming performance, whilst monitoring physiological response (heart rate). They found that players in the blue screen condition improved their performance with time and that players in the red screen the opposite effect was observed and their performance declined over time. This was associated with increased heart rate and it suggests that there is an arousal factor involved. This effect was not observed when varying the system sounds alone and a background colour change was required to observe the performance difference. They did not report on the effect of background colour change alone and so this confound prevents determining whether any arousal-caused performance changes were due to system sound or to background colour effects. Performance in this case may be linked to user preference because these findings suggest that users might prefer an environment that better suits their own arousal levels, which has implications for the design of user interface skins. These are now discussed in the following sections.

2.6.4 Linking Colour to Personality Trait

The previous section looked at the individual response to colour. The purpose of this section is to link colour to personality trait by examining research that has found links between user personality trait and user interface colour. Of the five factors in the Big Five model of personality one of the most researched is the extroversion trait and the research referenced here relates to this dimension.

Researchers have considered that personality and colour are linked and there are a number of studies to investigate this. Eysenck (1941) found that preference for modern paintings with bright colours were correlated with extraversion and older paintings with subdued colours were correlated with introversion (one of the dimensions described by Costa and McRae's Five Factor Model and measured by the IPIP-NEO inventory discussed previously in Section C). These findings were replicated by Barrett and Eaton (1947) who found that preference for bright and pure colours was correlated with extraversion and preference for tints and shades correlated with introversion. Eysenck's underlying hypothesis is that introverts are already highly aroused because they are more concerned with the internal processes of the mind and therefore seek to avoid additional external stimuli. Conversely, extroverts are less highly aroused internally and seek external stimuli. Therefore logically it follows that introverts may seek to avoid overly stimulating colours and seek more subtle colouring.

Research by Saati, Salem and Brinkman (2005) investigated relationships between user personality and user interface skin. They created two applications that featured five different coloured user interface skins that could be freely selected by the participant. They used the IPIP-NEO personality inventory to determine user personality and found correlations between the colour blue and introversion, with the blue interface skin selected more frequently than by extroverts. Their research was exploratory and considered a pilot test for this investigation.

Cerbus and Nichols (1963) found no significant relationship between colour and personality. However, they did not consider the context in which people prefer colour and Holmes and Buchanan (1984) found that colour preference was a function of the object in question and criticized previous research on the grounds that colours were

presented to participants in isolation of any application. Furthermore, the research conducted by Cerbus and Nichols (1963) pre-dates contemporary psychometric testing of the five factors and so the reliability of their personality data may be of questionable comparative value for this research.

With regard to user interface skins and colour selection, the research described above suggests that the personality dimension of extroversion can define a user segment for whom appropriate user interface skins may be provided.

2.6.5 Linking Interface Shape to Personality

The previous sections have been looking at colour as an interface factor that can be linked to user personality, to provide a more personally relevant interaction. This section now looks at interface shape as an interface factor and argues that like colour, interface shape can be linked to personality traits.

There is evidence to support a link between shape and emotional response. Kreitler and Kreitler (1972) found a tendency for respondents to rate the emotional response of drawn lines similarly: "sadness" depicted by lines going down and to the left or right, "gayness" by horizontal lines or lines going up and to the left and right, "anger" by irregular, jagged lines and "graveness" and "idleness" by gently curving or straight lines. Blank et al. (1984) suggest that these lines mirror postures associated with certain moods, for example droopy representing a posture associated with sadness. This evidence, although limited, does suggest that the variation of the shape of the user interface may also elicit an emotional response from the user, and may be a contributing factor for personally relevant user interface skin design. By producing an emotional response that is preferable to certain personality traits a more personally relevant interaction can be designed.

The shape of an object can be considered to be contributing to its visual complexity, defined by Martindale (1979) as the number of turns in a polygon. Berlyne (1957, 1958) and Berlyne and Lawrence (1964) found that when participants were looking at pairs of shapes, they spent longer looking at the more complex member of the pair. Berlyne suggests that complexity creates arousal potential in the user, and that this is represented in a Wundt curve (an inverted U function). This suggests that individuals seek out an optimal level of arousal, and it is this arousal modifying behaviour that

this research argues can be used to design a more personally relevant interaction. This finding also supports the Yerkes Dodson Law (Yerkes & Dodson, 1908) which predicts an inverted U-shaped function between arousal and performance. User arousal can be linked to user personality via user interface design because there is evidence to support arousal as a function of user personality (Eysenck, 1941; Barrett and Eaton, 1949). It should be noted for clarity that the arousal is not a primary concern for this research, but serves to inform potential underlying mechanisms responsible for differences in individual response as a function of user interface skin (e.g. arousal and skin shape complexity).

Studies by Martindale et al. (1979) were initially unable to replicate the findings of Berlyne. Their findings suggested that it was the perceived meaning abstracted from the shapes presented to participants that was responsible for arousal, and not the complexity of the shape as predicted. However, Martindale et al. (1990) did replicate the findings of Berlyne when they adjusted the experimental conditions (specifically the size of the room and the distance from participant to shape). The initial 1979 data that suggested that the shapes contained inherent meaning has to an extent confounded both Martindale et. al. (1979, 1990) and Berlyne (1979) because both experiments used the same shapes. This means that there was no control for meaning in these experiments. This provides a motivation for experimental hypothesis H_3, to test the effect of shape complexity on user behaviour. It is hypothesized that greater shape complexity (as defined by number of sides in a polygon) requires greater reticular activation and causes a greater arousal response. This is covered in greater detail in Chapter 3.

The Rorschach Ink Blot Test uses colour and shape to determine personality (Kunce and Tamkin, 1981). In the monochromatic versions of the test, shape is used exclusively to elicit responses that allow deduction of user personality. Participants are asked to abstract meaning from irregular shapes. This suggests that the shape of an object is partly responsible for eliciting different emotional responses. Norman (1988) describes affordances as the design aspect of an object which suggest how the object should be used. Like Rorschach ink blots, Norman's concept of affordances states that the form of an object implies something about its nature. In the case of ink

blots the shapes afford meaning, and it is the interpretation of this meaning that is responsible for the emotional response characterised by different personality types.

Martindale and Moore (1989) found that subject related meaningfulness accounted for 51% of the explained variance in preference, whereas subject rated complexity accounted for only 4%. It is possible that a similar interpretative phenomenon was influencing the Martindale and Berlyne studies, so this research also investigates meaning as an interface design feature that might be used to provide psychological customisation. The next section addresses this in greater detail.

2.6.6 Meaning and Similarity Attraction

The two previous sections have addressed interface features that influence user behaviour by operationalising arousal using colour and shape factors. This section argues that there are aesthetic factors relating to meaning within a skin, rather than the individual aesthetic factors (colour and shape) that may provide a means by which to design for personality traits. The meaning within a skin, for example any images or symbolic representation can be operationalised by representing personality within a user interface. Figure 2-13 below illustrates the differences between a skin that embodies a personality trait and one that does not (a non-meaning skin).

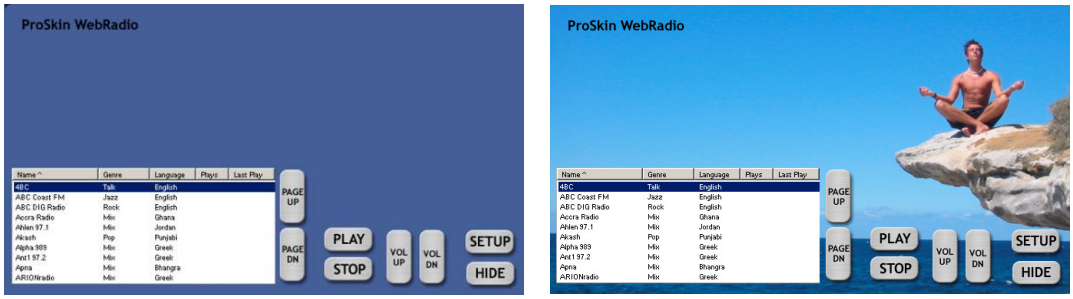


Figure 2-13 - Comparison between skins to illustrate the difference between a meaning skin and a non-meaning skin

The Similarity Attraction Hypothesis (Byrne and Nelson, 1965) states that perceived similarity predicts attraction, with people preferring people similar to themselves. Reeves and Nass (1996) found whilst testing the Similarity Attraction Hypothesis that users interact with computers in ways similar to other human beings. They referred to this as Computers As Social Actors (CASA). They found that computers that exhibited a dominant personality type were preferred by users of dominant personality

type, and vice versa for submissive personalities. With these studies in mind it seems logical that if a user interface skin were to represent a certain personality trait, users with similar personality traits would prefer it over other personality traits. This research differs from that of Reeves and Nass in that their research focussed on the tone and mediators of interaction whereas this research focuses on the effect of visual appearance on personal preference.

Van Montfort (2006) has investigated the effect of gist and meaning in images. She found that the meaning interpreted from pictures (gist) was more important than colour changes within the content in accurately perceiving meaning. This was because users noticed minor changes to the images that changed the gist of the entire image, but that they did not notice colour changes (for example hair colour). This research suggests that the meaning conveyed by component parts of an image is more important to the successful interpretation of the gist of the image than colour is.

The effect of meaning on user behaviour is tested by hypothesis H_4 (users prefer user interface skins that exhibit a similar personality to their own) and can be found in greater detail in Chapter 4.

2.7 Chapter Review and Experimental Hypotheses

This chapter has argued that the separable user interface architectures exist that facilitate customisation and change of the user interface. Furthermore it has argued that profiles may be created from segments defined by user personality, derived from implicit data collection. It also has argued that interaction may happen as a function of user personality, and that in order to provide a more personally relevant interaction, the user interface can support design for personality traits through visual aesthetic factors such as colour, shape and meaning.

The experimental hypotheses for this research will be tested in the following chapters. They are summarised in Figure 2-14 overleaf:

EXPERIMENTAL HYPOTHESES

H_1 Personality data can be extracted from log files
H_2 User personality can predict preference for user interface skin colour
H_3 User personality can predict preference for user interface shape
H_4 Users prefer user interface skins that exhibit a similar personality to their own

Overall Research Motivation:
A more personally relevant interaction can be produced using profiled user interface skins

Figure 2-14 - Experimental hypotheses

Chapter 3 – METHODOLOGY

3.1 Overview

This chapter describes the design and details of the experiments conducted during the investigation of the experimental hypotheses. The chapter starts out by describing the research approach taken for each of the four hypotheses. The chapter then continues by providing details of the experimental approach, including attracting participants, procedure, materials and measures. A limitations section describes the known limitations of the approach and a summary section concludes the chapter. Details regarding the experimental platform are provided in the next chapter, chapter 4. A visual guide to this chapter is provided below in Figure 3-1:

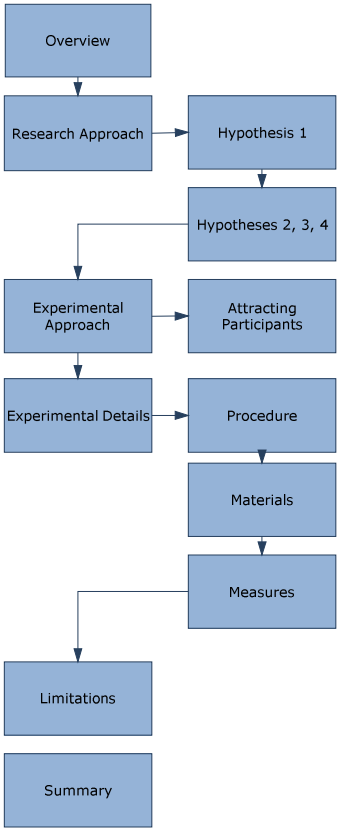


Figure 3-1 - Visual guide to chapter 3

3.2 Research Approach

This section describes the research approach taken to investigate the experimental hypotheses. The overall research motivation of this research is that a more personally relevant interaction can be produced using profiled user interface skins. From this, four experimental hypotheses have been proposed in Chapter 2. In order to investigate these four experimental hypotheses two different research approaches have been taken. This is because the experimental hypotheses can be divided into two groups, each with an appropriate approach (illustrated below in Figure 3-2). Hypothesis 1 aims to establish a behavioural measure of personality. Hypotheses 2, 3 and 4 are attempts to identify user interface design factors that can be linked to the personality traits established in Hypothesis 1.

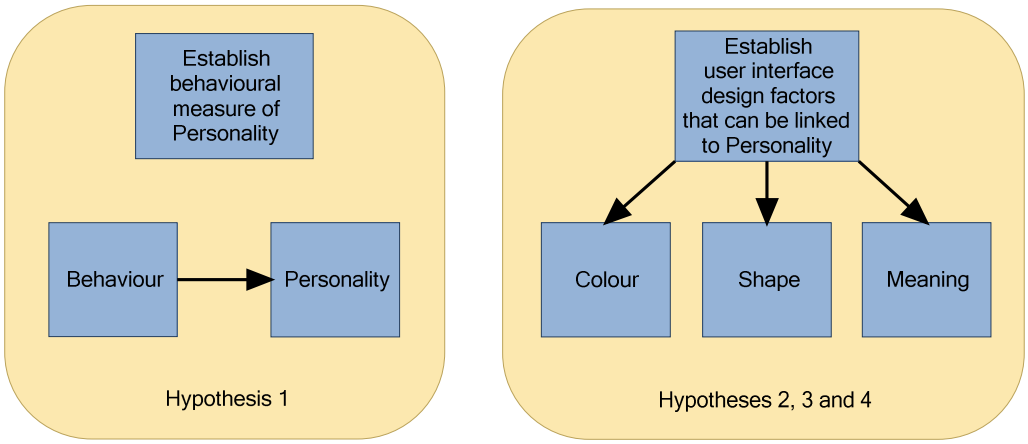


Figure 3-2 - Two different research approaches to address the four experimental hypotheses

3.3 Hypothesis 1: Establishing a Behavioural Measure of Personality

The approach taken to investigate the first experimental hypothesis is to compare interactive behaviours captured and recorded in log files with participant responses to an established personality inventory, the IPIP-NEO. By looking for correlations between interactive behaviour and personality trait scores obtained from questionnaires, a behavioural measure of personality may be produced enabling

personality to be determined from interactive behaviour and without the need to explicitly involve the participant.

3.4 Hypotheses 2, 3 and 4: Establishing User Interface Design Factors to link to Personality

The approach taken to investigate the remaining three experimental hypotheses is to select three user interface design factors to be manipulated in order to determine the effect on skin preference as a function of personality trait: colour, shape and perceived meaning. Correlations found between personality trait scores and variations in each of these experimental conditions (e.g. different colours or shapes) provide a means to link user interface design factors to the personality of the participant.

3.5 Conceptual Model, Variables and Operationalisation

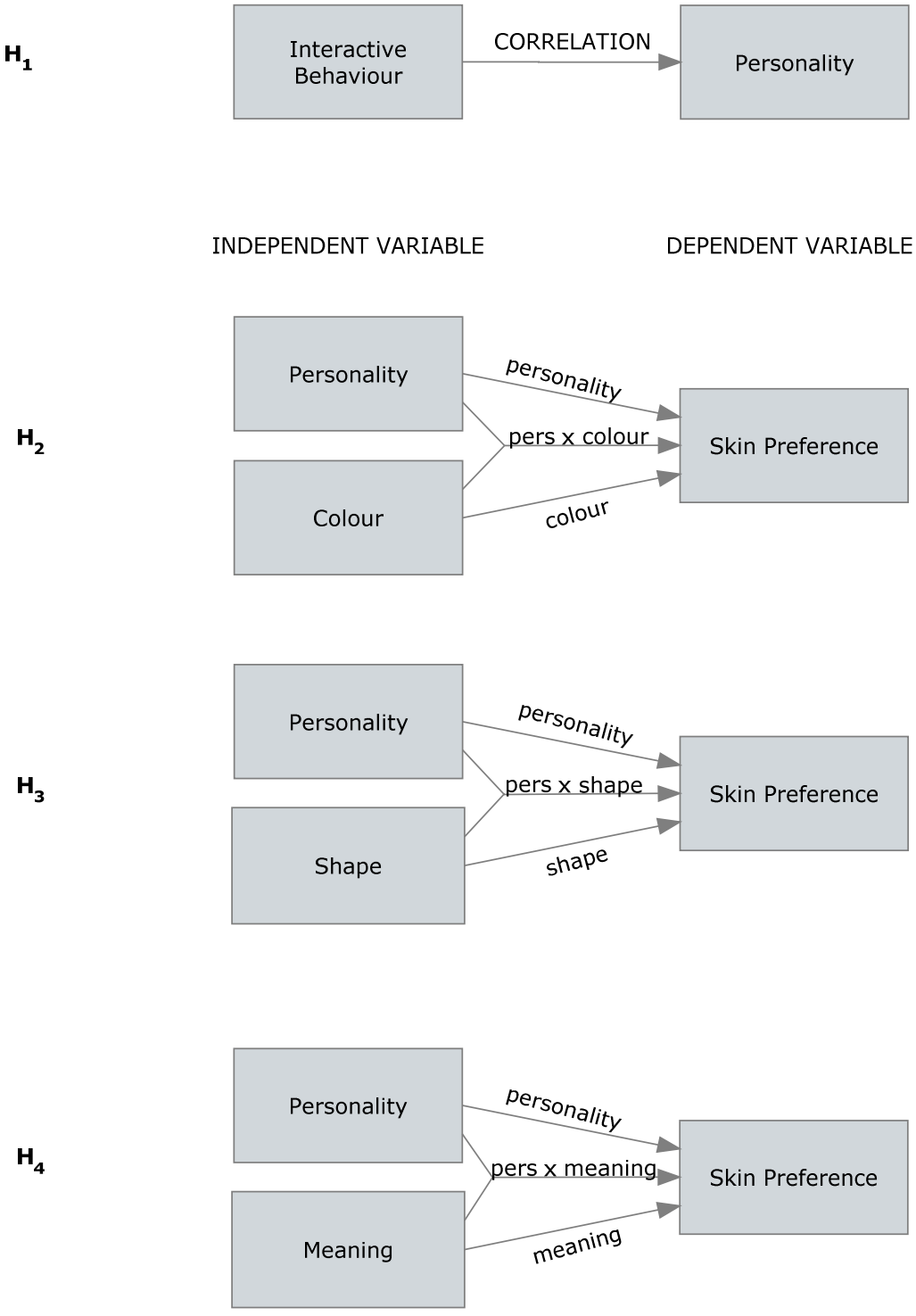


Figure 3-3 - Table summarising the experimental variables for each of the four hypotheses

Figure 3-3 illustrates relationships between the abstract constructs that were studied. For H_1 the independent variable is interaction behaviour of the user and the dependent variable the personality of the user. The abstract construct interaction behaviour was operationalised in this study by a series of behavioural measures (Figure 3-4) that were recorded in a log file, while personality was measured by applying the IPIP-NEO personality inventory. Figure 3-3 also shows the conceptual model underlying the other three hypotheses. Although the independent variables are different for all three models, each has skin preference as the dependent variable. The user's preference towards a skin was operationalised across two dimensions: (1) an objective dimension referring to the user's behaviour, and (2) a subjective dimension referring to the users stated preference. The objective dimension was operationalised with one indicator, which was ratio of the time a user used application with the specific skin and time the user used the application. The subjective dimension was operationalized by the following four Likert type statements on which users rate their likelihood that would use a specific skin in four different situations:

- "I would select this skin when working on my PC"
- "I would select this skin when working on my own"
- "I would select this skin when I am not alone"
- "I would select this skin when I am using the PC recreationally (e.g. family photos, playing games, using the WWW)"

For H_2 the independent variable was skin colour and user personality. The latter was again operationalised with the IPIP-NEO inventory. The skin colour was operationalised as an eight level nominal variable, implemented as eight skins with the colour red, yellow, green, blue, pink, orange, purple and grey. The hypothesis was studied by examining what effect the combination of skin colour and user personality had on user preference of a skin. For H_3 the independent variables were meaning and user personality, with personality operationalised with the IPIP-NEO inventory. Meaning was operationalised as the perceived personality of images within eight user interface skins, measured using the TIPI personality inventory. The hypothesis was studied by examining what effect the perceived personality of the user interface skin had on the user's skin preference, as a function of the user's personality.

Hypothesis	Independent Variables		Dependent Variables	
H_1	**Interactive Behaviour** defined as Total Clicks Session Time Station Change Volume Change Clicks Per Second Extra Effort (Mean) Extra Effort (Sum)		**Personality Data** Personality data from IPIP-NEO personality inventory (120 item psychometric) values: Openness Conscientiousness Extroversion Agreeableness Neuroticism	
H_2	**Colour** defined as Hue, Saturation, Light values: (Hue = X Sat = 50% Light = 0%)	**Personality** measured by IPIP-NEO 120 item psychometric values: Openness Conscientious. Extroversion Agreeableness Neuroticism	**Skin Preference** measured subjectively and objectively: SUBJECTIVE MEASURE Skin Preference Questionnaire measured as 4 item 7 point Likert statements	OBJECTIVE MEASURE Behavioural Measures defined as time skin X Total time skin pack
H_3	**Shape** defined as number of turns in shape values: 5,7,10,13,20 30,40	**Personality** measured by IPIP-NEO 120 item psychometric values: Openness Conscientious. Extroversion Agreeableness Neuroticism	**Skin Preference** measured subjectively and objectively: SUBJECTIVE MEASURE Skin Preference Questionnaire measured as 4 item 7 point Likert statements	OBJECTIVE MEASURE Behavioural Measures defined as time skin X Total time skin pack
H_4	**Meaning** measured by TIPI 10 item psychometric values: Openness Conscientious. Extroversion Agreeableness Emotional Stability	**Personality** measured IPIP-NEO 120 item psychometric values: Openness Conscientious. Extroversion Agreeableness Neuroticism	**Skin Preference** measured subjectively and objectively: SUBJECTIVE MEASURE Skin Preference Questionnaire measured as 4 item 7 point Likert statements	OBJECTIVE MEASURE Behavioural Measures defined as time skin X Total time skin pack

Figure 3-4 - Table summarising the experimental variables for each of the four hypotheses and how they were operationalised into indicators

3.6 Applying Research Approaches within ProSkin

Both research approaches are accommodated within a single experimental platform. In order to test the experimental hypotheses, a methodology was required that facilitated collection of skin rating data, personality data and interactive behaviour data from participants. The prevalence and availability of both processing and networking resources provide an infrastructure for the collection of participant data, and technologies such as the personal computer and "always on" Internet connections (e.g. broadband services) mean that participant data can be collected and transmitted to the experimenter automatically. By removing the need for the experimenter and participant to be physically co-located, computer supported data collection of this type may offer experimental HCI useful resource savings (e.g. reduced time, increased participant numbers). In addition to any potential resources saved by this type of methodology it may provide a means to observe and record natural user behaviours in ways that cannot be achieved in the controlled environment of an observation lounge.

3.7 Experimental Approach Overview

In Chapter 1 the vision for this research was described. To briefly summarise, the vision is of large number of participants contributing non-personally identifiable, low effort usage data to a central data repository. In order to achieve this vision a large number of participants were required, and in order to achieve that a highly visible campaign to attract participants to the study was required. The next section discusses the experimental approach adopted and the reasoning behing the decision. The next sections will describe in detail the methods used to attract participants to the study and then consider the implications for the experimental design.

3.8 Experimental Approach

The vision for this research involves large numbers of participants to contribute data and requires high participation to generate experimental data for analysis. In order to

investigate the four experimental hypotheses it was decided that a single sample approach would be most appropriate. This meant investigating all four experimental hypotheses (i.e. both research approaches) in a single data capture period. This was because in order to make an impression with potential participants, a single campaign would create awareness and direct people to the project website to participate, instead of having separate sampling periods for each hypothesis with an associated campaign, each perhaps having a lessening impact on potential participants. A disadvantage of using a large single sample as opposed to four separate samples is that it provides the opportunity to revise the experimental design or fix any practical problems that might arise at each step of the investigation. As a result of this an extended beta test period was incorporated into the design cycle in the development of the experimental platform, so that as many procedural and technical problems could be tested and fixed prior to the commencement of the single sampling period. The beta test period for this research was three months in total from December 2006 to February 2007, when the full experimental platform was tested for functionality, security, scaleability, useability and stability prior to the start of the experiment in August 2007. The beta test was comprised of a total of 31 participants, of which 12 were male (mean age 28.9, standard deviation 6.5) and 19 were female (mean age 30.6, standard devision 5.6).

3.9 Attracting Participants

In order to attract a large number of participants on a limited financial budget, a number of different strategies were used in order to maximise awareness and drive participation across a number of diverse media:

Postcard Campaign: 3000 postcards were produced that advertised the experiment. They were placed on noticeboards all around Brunel University, and in free postcard racks in as many Central London underground stations, cinemas, bars and restaurants as possible. This activity was started just prior to the commencement of the live experiment and continued throughout the sampling period.

Press Release: Brunel University's public relations department issued a press release giving details of the experiment and encouraging people to participate.

Poster Campaign: In addition to the postcards, A4 posters were produced and placed around Brunel University. Event-specific posters were produced for the Ashes cricket matches that took place during the sampling period, because radio is a popular medium for listening to cricket matches and provided an appropriate opportunity to attract users.

Internet Radio DJ: The experimenter became an Internet radio DJ for a period of three months during the sampling period to promote participation to radio users. Two radio stations were broadcast from, one called UK Nova Radio (www.uknova.com) that is the radio station for a website that provides legal non-commercial television programmes on the Internet, and another called Pure Radio, an in-game radio station for the massively multiplayer online roleplaying game Eve Online (www.eveonline.com). DJ appearances were more frequent for Pure Radio than for UK Nova, with broadcasts at least three times a week during the sampling period, most frequently 10pm until 1am GMT. UK Nova broadcasts were once a week on Sundays between 6pm and 8pm GMT. The reason for appearing on two different radio stations was to attempt to vary the listenership attracted to participate because Eve Online demographics are skewed towards males between the ages of 18 and 30, whereas the UK Nova demographics are more likely a much broader range of listeners of both sexes. There are no statistics available as to the demographics of either listener group, although anecdotal evidence suggests that the Eve Online demographic is heavily skewed towards males between 18 and 30.

E-mail Campaign: Approximately 250 e-mails were sent to individuals and mailing lists, combining both academic institutions and corporate entitites (e.g. Brunel University, Hays Legal, ACM, Usability News). The e-mails to individuals were personal friends or professional acquaintances of the experimenter.

3.10 Experimental Details

This section details how the two research approaches described in section 4.2 were applied to investigate the four experimental hypotheses.

3.10.1 Procedure

Participants downloaded an Internet radio player application that had been specifically written for this research from the project website (www.proskin.org) and installed it on their own computer. When the application was run for the first time, users were automatically allocated a unique ID number and randomly assigned to one of the twelve experimental conditions (see

Condition	Skin Pack 1	Skin Pack 2	Skin Pack 3	Skin Access
1	Colour	Shape	Meaning	Delay
2	Colour	Meaning	Shape	Delay
3	Shape	Colour	Meaning	Delay
4	Shape	Meaning	Colour	Delay
5	Meaning	Shape	Colour	Delay
6	Meaning	Colour	Shape	Delay
7	Colour	Shape	Meaning	No Delay
8	Colour	Meaning	Shape	No Delay
9	Shape	Colour	Meaning	No Delay
10	Shape	Meaning	Colour	No Delay
11	Meaning	Shape	Colour	No Delay
12	Meaning	Colour	Shape	No Delay

Table 3-5 on the next page). All users received the default and helpskin at installation, and then received each of the three skin packs (colour, shape, meaning). The order that users received skin packs in was controlled for, creating six conditions. A week long delay condition was designed for, extending the total number of conditions to twelve, although during the data analysis this distinction was not considered. Users initially received three questionnaires to complete, with the General Information and Skin Preference questionnaires mandatory for completion before users received new user interface skins. Full details of the progression criteria can be

found in section 4.13.1. Once the progression criteria had been satisfied users received a skin pack that was downloaded directly to their computer, according to the experimental condition. The user then received additional skin packs every three days until all skin packs had been received, at which time server communication was halted and all experimental skins made available for the user.

Condition	Skin Pack 1	Skin Pack 2	Skin Pack 3	Skin Access
1	Colour	Shape	Meaning	Delay
2	Colour	Meaning	Shape	Delay
3	Shape	Colour	Meaning	Delay
4	Shape	Meaning	Colour	Delay
5	Meaning	Shape	Colour	Delay
6	Meaning	Colour	Shape	Delay
7	Colour	Shape	Meaning	No Delay
8	Colour	Meaning	Shape	No Delay
9	Shape	Colour	Meaning	No Delay
10	Shape	Meaning	Colour	No Delay
11	Meaning	Shape	Colour	No Delay
12	Meaning	Colour	Shape	No Delay

Table 3-5 – Experimental conditions controlling for skin pack order

3.10.2 Materials

The only materials used were to promote the research and to attract participants. These consisted of postcards and posters, the former professionally printed (www.printing.com) and the latter printed by the experimenter. One of the postcards is shown below in Figures 3-6 and 3-7:

Figures 3-6 and 3-7 - Promotional postcard used to promote awareness and use of the WebRadio, showing front and back.

3.11 Limitations of Method

Using an externalised methodology where the experimenter is remotely situated to the participant has inherent limitations due to the environmental variables alone (Birnbaum, 2000). In such a non co-located situation there will be environmental variation that either directly or indirectly influences participant behaviour (e.g. speakers versus headphones, interruptions, ambient noise etc.). Whilst on one hand this can be considered a limitation, on the other it may be considered by some as an advantage, providing a wholistic, "real world" account of behaviour in a natural context of use. Such uncontrolled environmental variables reinforce the need for sufficiently high participation contributing valid data, so that the effect of uncontrolled variables is minimised.

In addition to the externalised methodology, a one sample approach has been taken. As was stated earlier, the reasoning for this was to take advantage of a single large marketing effort to attract participants to the project. This was because it was felt that subsequent marketing efforts would have diminished effect on potential participants. As a result of this it meant that any methodological issues that arose during the course of the four experiments could not be rectified in subsequent iterations. An extended beta testing period of three months allowed the platform and methodology to be tested on a small scale of 30 users (who were not allowed to participate in the main experiment), although such a testing period with small numbers of participants cannot show all the potential issues.

In comparison to studies where the experimenter and participant are physically co-located (for example in an observation lounge), externalised research such as this is unable to moderate as effectively the types of participants that are accepted. For example, in the example of an observation lounge study the experimenter might be waiting in a corridor to ask passing people if they would participate. At this point the experimenter is able to select males or females to ensure a balance of gender in their experimental conditions, or only select people that look over the age of 18. In an externalised research methodology such as this, the filtering of participants typically performed by the experimenter is performed by the computer (i.e. the experimental platform). As for the wholistic view of externalised methodologies stated at the

beginning of this section, the apparent difference in the selection and acceptance of participants between physically co-located and externalised methodologies may be considered both a limitation and a strength. On one hand, the removal of the experimenter from the selection process removes a human variable and certain experimenter bias, and on the other hand leaving selection to a non-human computer allows for potentially inconsistent data from untruthful participants (e.g. age under 18).

3.12 Summary

This chapter has provided the details of the experimental methodology and approach. The next chapter describes the details of the experimental platform used to investigate the experimental hypotheses.

Chapter 4 – EXPERIMENTAL PLATFORM

4.1 Methodology

This chapter provides the details of the experimental platform developed to support the investigation. The four chapters following this one (5 to 8) will each provide hypothesis-specific experimental detail so that each hypothesis is handled within its own chapter.

H_1	Chapter 4	Personality data can be extracted from log files
H_2	Chapter 5	User personality can predict preference for user interface skin colour
H_3	Chapter 6	User personality can predict preference for user interface shape
H_4	Chapter 7	Users prefer user interface skins that exhibit a similar personality to their own

Table 1 - Hypothesis-specific chapters

A visual guide to the chapter can be found in Figure 4-1.

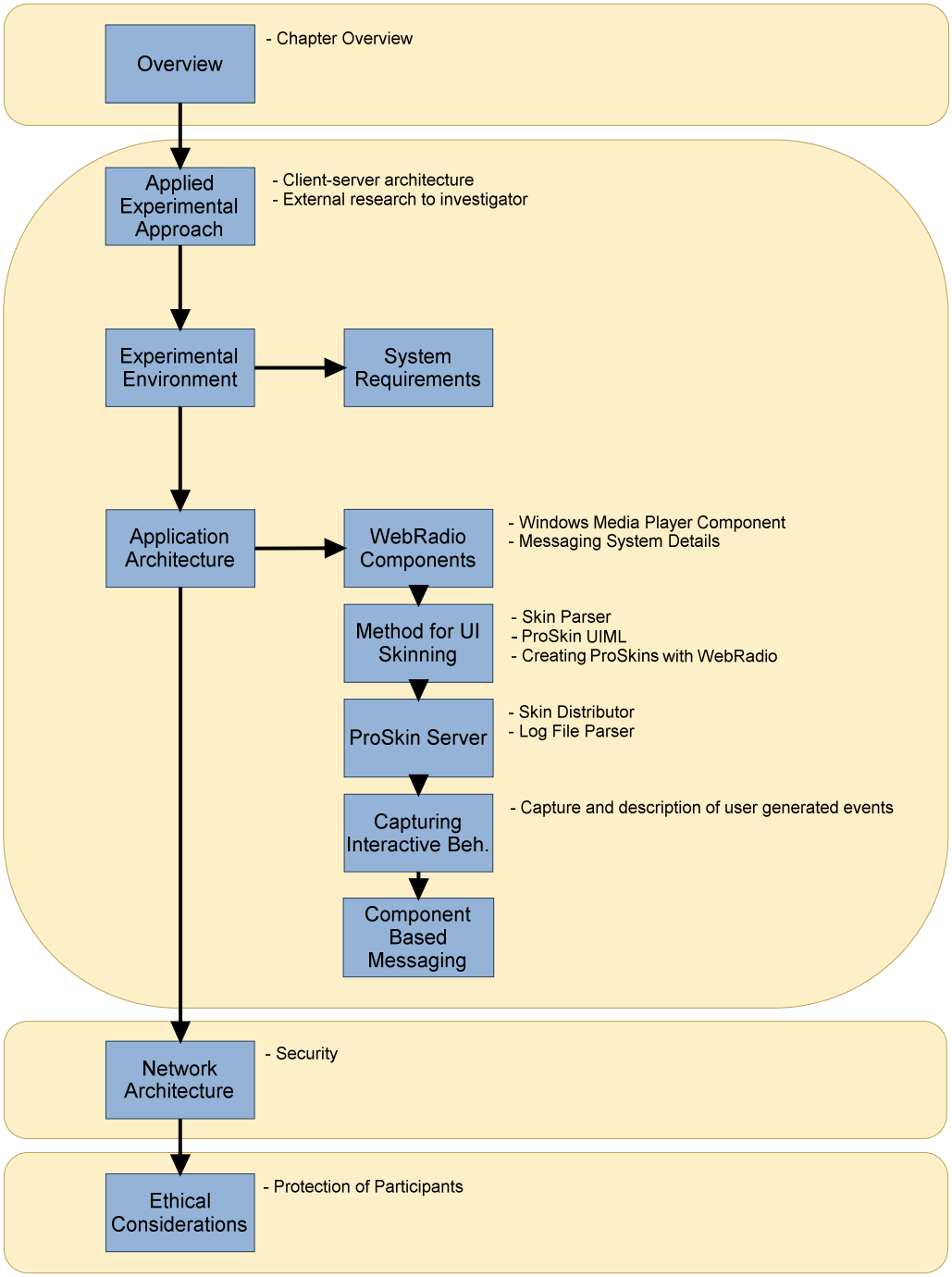


Figure 4-1 - Visual guide to Chapter 4

4.2 Applied Experimental Approach

A client Internet radio application, "WebRadio", was developed that featured changeable user interface skins, log file recording of interactive behaviour and the

administration of questionnaires, so that both research approaches could be accommodated within a single platform and the experimental hypotheses tested. The WebRadio was distributed to non-local participants (via the project web site) along with different user interface skins and log files of interactive behaviour and questionnaire responses collected stored for analysis on the server in the database . This is illustrated below in Figure 4-2. By looking for correlations between personality trait scores (collected with the IPIP-NEO questionnaire) and interactive behaviours relating to user interface factors (e.g. colour, shape, meaning, self reported skin preference questionnaire) the experimental hypotheses can be tested.

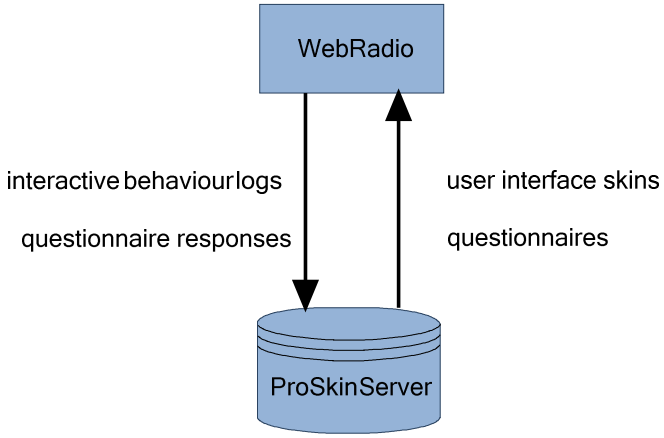


Figure 4-2 - Communications between Client and Server

Having described the experimental approach, the next section addresses the application of the approach starting out with the system requirements. After the requirements are detailed there follows an overview of the actual experimental platform (and systems) and then the rationale and detail for the architecture and construction.

4.3 Experimental Environment

The purpose of this section is to describe the experimental environment that was developed for this research. At the outset the requirements of the experimental platform are listed. Following this there is a brief discussion on the selection of an experimental application type and then the section moves into describe and justify the design and architecture of the platform created in response to the requirements.

4.4 System Requirements

In order to investigate the experimental hypotheses a number of features were required of the experimental platform:

1) Changeable user interface: an application featuring the ability to change the user interface quickly and easily by end users.
2) Popular application type: an application that would have broad usage appeal in order to elicit regular unprompted usage from participants to facilitate the generation of user data.
3) No personal content: an application that did not include or require the use of participant content (e.g. MP3s, text documents, WWW pages) to protect participants and to encourage usage.
4) Moderate complexity of interaction: an application that featured sufficient opportunity to elicit interactive behaviours without being overly complex.
5) Large target participant audience: the application should be able to be run by a majority of personal computer users in the world to maximise the opportunity to collect user data.
6) User interface skins: development of user interface skins for each of the experimental hypotheses (colour, shape, perceived personality).
7) Skin distribution: a means to control deployment of ProSkins to participants according to the experimental conditions
8) Log File Recording: a means to capture user data, both questionnaires and interaction data relating to the user interface skin.
9) Internet communications: a means to distribute experimental skins and to capture log files from remote clients via the Internet.
10) Questionnaire administration: the ability to distribute user questionnaires and collate the responses.
11) Communication with participants: a means to contact participants to inform them of new questionnaires to complete and to provide a feedback mechanism to experimenters.
12) Ethical handling of participant data: protection of participants and their data through system supported anonymity and withdrawal from experiment.
13) Multiple contiguous user support: the ability for multiple users to access the system simultaneously without data loss.

The next sections address the platform built to accommodate these requirements, starting out with a short discussion on the selection of an application type before

providing detail of the experimental platform and the measures taken to meet these requirements.

4.5 Selection of an Application Type

A number of different application types were considered in the design and planning phase: an Internet browser, a text editor, a calculator and a digital music player. The application would need to provide a relatively straightforward interaction, defined as having limited interface components for interaction (e.g. buttons). This was because the interface needed to offer enough interactive variance to allow a variety of interactive behaviours to be observed but not so many as to provide an overtly complex interaction to analyse. It was for this reason that both the text editor (e.g. Microsoft Notepad) and calculator (e.g. a non-scientific calculator, such as the calculator in Windows 2000/XP) were discounted as viable experimental applications because they both offered too simple an interaction that did not appear to offer sufficient breadth of interaction to be of use in this research. The Internet browser was discounted because it appeared to offer too broad an opportunity to interact because of the individual nature of website design. This was because each website provides a different interaction as a function of the content and presentation of the site itself and therefore was considered to not provide enough experimental control for the purposes of this work.

A digital music player was selected for development because it could offer a limited interaction but without being too simple an interaction, such as the calculator. This meets requirements 2 (popular application) and 4 (appropriate complexity) of the system requirements as listed on page 60. In order to avoid any legal problems, an Internet radio player (WebRadio) was selected rather than a digital music file player (e.g. MP3, WAV files). This was because Internet radio stations are broadcast, removing any potential user concerns arising from the logging of user music listening behaviours, because there is no logging of locally held user content. This also meets the system requirement 3, to not include or require the use of participant content.

The following sections of this chapter will describe in detail the experimental platform constructed, including the application and network architectures developed.

4.6 Experimental Environment Overview

The ProSkin experimental environment is a client-server architecture developed using the Microsoft .NET programming language and Microsoft Access database, the client being the WebRadio application and the server referring to the database and collection of web services for the management of the experiment and collection of data. An overview of the application architecture showing the client server relationship is presented in Figure 4-3 overleaf. It provides an overview of the application architecture from a top down perspective with the user and user interface skin placed at the top of the image. The image also illustrates the distributed nature of the architecture with the Internet providing wide area communications between components.

The image represents the WebRadio client and the ProSkin Server and shows their internal components and intercommunication, both internally and across external networks (i.e. the Internet). The system functionality is provided through the intercommunication between these components, for example the messaging or skinning functions. Such functionality is referred to as a subsystem by this research and these subsystems are isolated and explained in further detail in the subsequent sections.

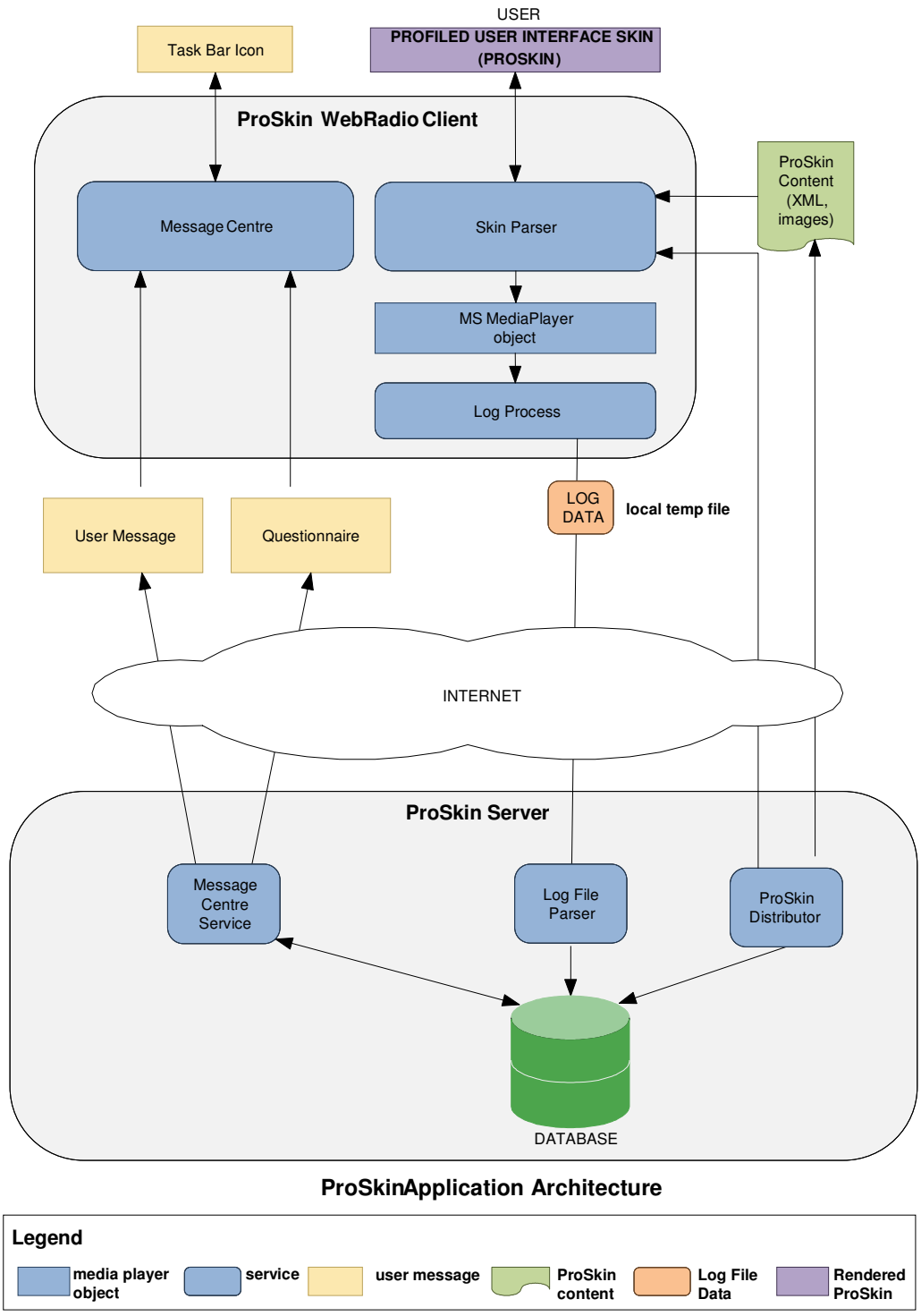


Figure 4-3 - ProSkin Application Architecture, showing client-server relationship

The platform was developed using Microsoft technology in order to meet system requirement 5, that the client application should be able to be run on a majority of personal computers in order to maximise potential participation. The Microsoft

Windows XP operating system is popular and widely installed globally, meeting the system requirement.

The first section will first look at the application architecture and describe the construction of the WebRadio client and all the component parts, and then the next sections will look at the ProSkin server and network architecture to support the application, as well as the ethical issues accommodated by the platform.

4.7 WebRadio Client Architecture

An Internet radio station player, named "WebRadio" was developed to meet the experimental requirements outlined in Section 3.8. The WebRadio application is comprised of four main subsystems:

- Microsoft Windows MediaPlayer object: a Microsoft programming component (an ActiveX control encapsulated within a .NET wrapper) providing the media player functionality (e.g. playing Internet radio streams, volume controls, station directory controls).
- Skin Parser: a function to parse incoming ProSkins and to render them for use as user interface skins for the MediaPlayer controls.
- Message Centre: a function to display messages and questionnaires to participants
- Log Process: a function to automatically record interaction with the Skin Parser to a locally held log file, and then to transfer that log file to the server at the end of each session.

An architectural diagram showing these component parts and their relationships to one another is shown below in Figure 4-4. The WebRadio client represented below is shown in the next section in the context of entire application architecture (both client and server) but is shown here to illustrate the elements comprising the WebRadio application. Further information regarding the construction of each of these components now follows, each component handled within its own section.

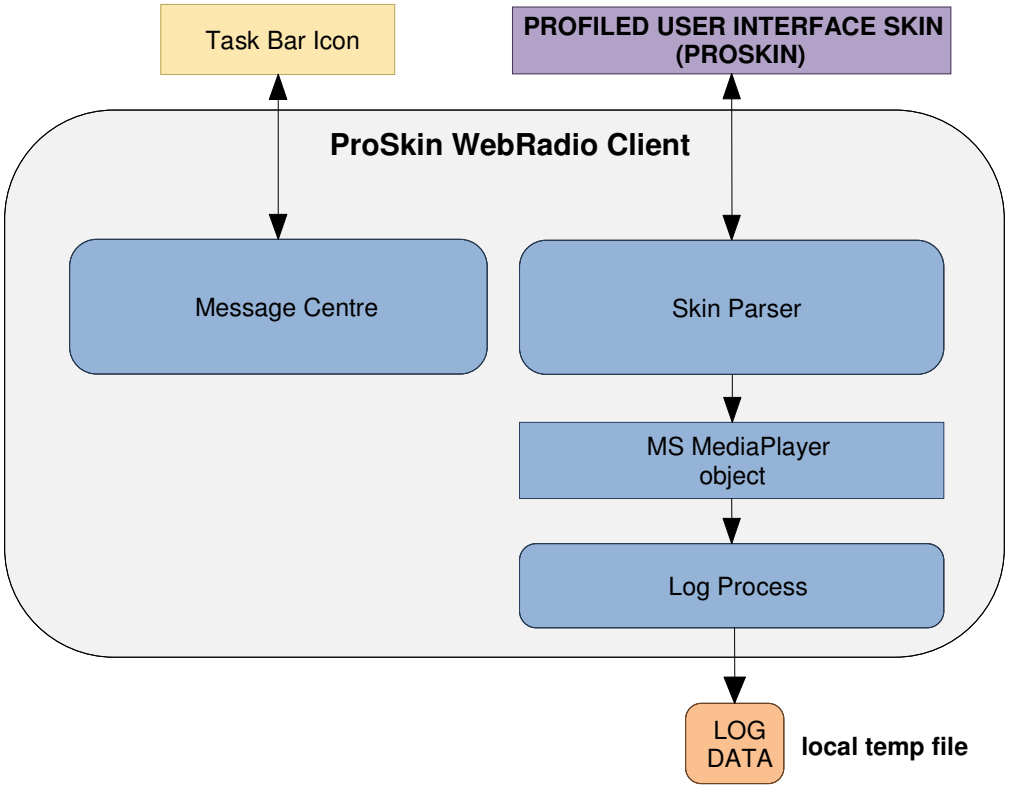


Figure 4-4 - Client Architecture (WebRadio)

4.8 WebRadio Client Components

4.8.1 Microsoft Windows Media Player Object

The Microsoft Windows Media Player object is a freely available programming construct that provides the media player functionality and programming controls (for example the connection to, decoding of and control of an Internet radio stream (e.g. stop, play) and user controls such as volume. It is made available by Microsoft in the form of the Windows Media Player Software Development Kit (SDK).

4.8.2 Message Centre

This section describes the Message Centre function within the client. At the outset the functionality is described along with the design decisions taken during the development. The section then continues to describe the different message types used to communicate with participants and the architecture to support it.

4.8.2.1 *Design and Functionality of the Message Centre*

The main function of the Message Centre is to provide a means to communicate with participants. This is so that both questionnaires and informational messages can be transmitted and received to and from individual clients (participants). The Message Centre handling of participant messaging and questionnaire administration meets the system requirements 10 (questionnaires) and 11 (communication) as described in the list of system requirements in Section 3.8. It was decided that message centre functionality, involving the delivery of informational messages and questionnaires, should be presented to the user independently of the WebRadio user interface. This meant that the user interface skins for the WebRadio were not required to support additional messaging functionality and could therefore be designed to support interactive behaviours relating exclusively to the radio player functionality.

In order to be able to provide a messaging function, the Message Centre function is implemented separately to the WebRadio interface via a Windows taskbar icon, shown below in Figure 4-5(a). The user right-clicks on the WebRadio taskbar icon and is presented with a context menu, shown in Figure 4-5(b). Clicking on 'View Messages' opens the Message Centre, shown in Figure 4-5(c).

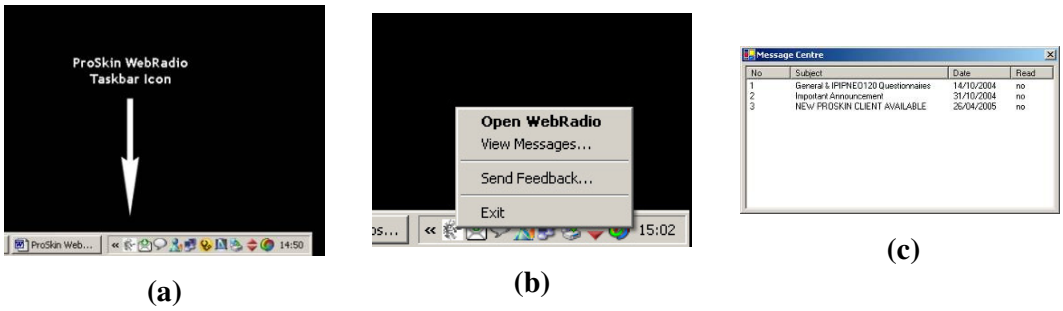


Figure 4-5 - The Message Centre function is accessed via the Windows taskbar icon (a), a context menu provides the access to view messages (b) which then opens the MessageCentre (c) when clicked.

The next section addresses the different kinds of user messages that are handled by the Message Centre, and the different means by which they are displayed to the user.

4.9 Message Types

There are two main types of message sent from the ProSkin server to the Message Centre in the WebRadio client, User Messages and Questionnaires.

4.9.1 User Messages

User Messages are simple text messages of up to 255 characters. Their main function is to allow communication from the experimenter to all participants or to specific individuals. User messages of this kind are not directly related to the primary hypothesis testing functionality of the platform but provide an important secondary function of communicating with remotely situated participants. They are displayed in a small pop-up window that spawns in the centre of the screen (illustrated below in Figure 4-6).

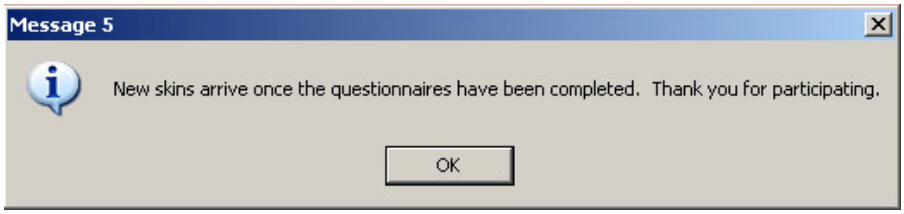


Figure 4-6 - Example of a User Message

4.9.2 Questionnaire Messages

Questionnaire messages are messages that link directly to a hypertext markup language (HTML) questionnaire hosted on the ProSkin world wide web (WWW) server. Questionnaire messages appear the same as User Messages to the user in the Message Centre but instead of opening a simple message window they open a limited WWW browser window to display a universal resource locator (URL). All questionnaires are then presented to the user as HTML forms generated by the ProSkin web server. The questionnaire was pilot tested during a pilot test of the entire system (n=46) and captured data as intended without any negative user feedback. An example of a Questionnaire Message is shown below in Figure 4-7:

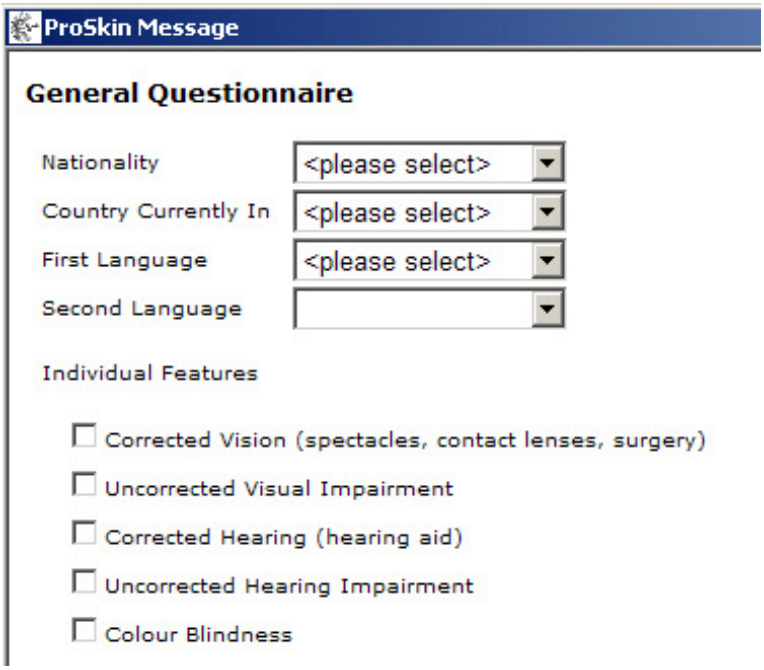


Figure 4-7 - Example of a Questionnaire message taken from the General questionnaire

4.10 Messaging Subsystem Architecture

This section describes the mechanics of messaging between client and server. First of all an overview of the messaging subsystem architecture is illustrated in Figure 4-8 below. It is represented in the overall application architecture in Figure 3:4 nested on the left hand side within the context of the other subsystems and components. The representation overleaf in Figure 3:9 provides additional detail, showing the two different types of message communicated between client and server (info and web, discussed in Section 3:13).

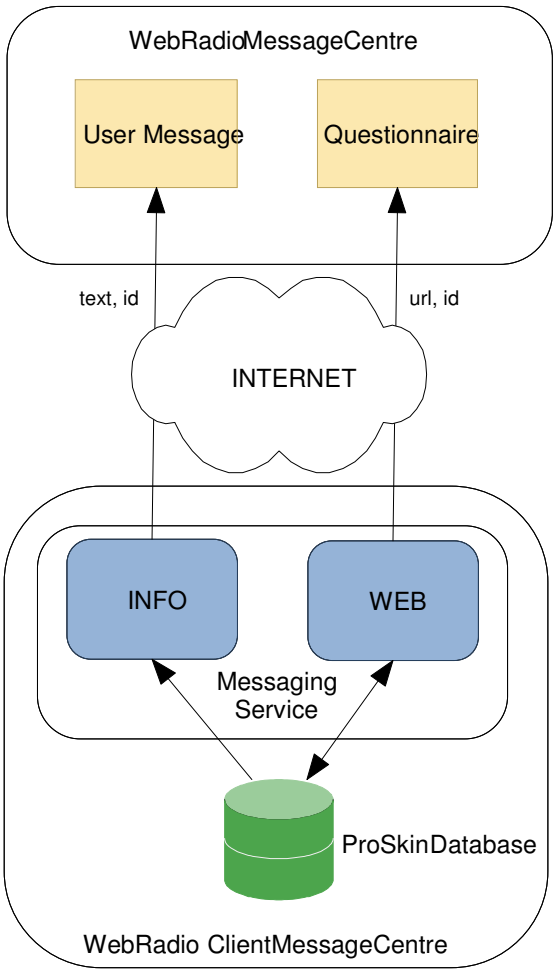


Figure 4-8 - Messaging architecture showing client server relationship

The illustration above represents the communication of messages between the ProSkin database on the server and the WebRadio client running on the participant's computer. When the WebRadio client is initialized there is a check for new messages from the database (the "user_messages" table). If there are new messages available they are sent directly to WebRadio client. An extract from the messaging table of the database is shown below in Figure 4-9 to show the two different types of messages.

mc_id	mc_type_id	mc_msg_date	mc_data	mc_msg_subject	mc_recipient
29	2	25/04/2006	http://server.proskin.org/general/general.aspx	General Questionnaire	ALL
30	2	25/04/2006	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire	ALL
32	2	12/08/2006	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire	ALL
33	2	12/08/2006	http://server.proskin.org/stomp/stomp.aspx	Stomp Questionnaire	ALL
39	2	13/10/2006 17:04:49	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	44E8A7DA
40	1	16/10/2006	New skins arrive once the questionnaires have been c	Complete Questionnaires to get new skins!	ALL
41	1	16/10/2006	Successful launch and WebRadio now live. Please in	Update 16/10 : Welcome to ProSkin WebR	ALL
42	2	16/10/2006 21:36:01	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	1AE71B91
43	2	16/10/2006 23:25:23	http://server.proskin.org/general/general.aspx	General Questionnaire (reminder)	5CAC5772
44	2	16/10/2006 23:25:51	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	5CAC5772
45	2	19/10/2006 15:56:05	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	64CED752
46	2	19/10/2006 15:56:28	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	64CED752
47	2	19/10/2006 23:20:17	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	1AE71B91
48	2	25/10/2006 09:55:28	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	9C423A42
49	2	01/11/2006 12:49:29	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	0666FBFF
50	2	03/11/2006 14:03:58	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	9C423A42
51	2	05/11/2006 23:22:12	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	085219D4
52	2	19/11/2006 18:55:59	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	1A0C183D
53	2	19/11/2006 19:00:37	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	1A0C183D
54	2	19/11/2006 20:31:28	http://server.proskin.org/stomp/stomp.aspx	Stomp Questionnaire (reminder)	1A0C183D
55	2	25/11/2006 14:11:28	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	085219D4
56	2	27/11/2006 17:42:36	http://server.proskin.org/general/general.aspx	General Questionnaire (reminder)	085219D4
57	2	27/11/2006 17:42:53	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	085219D4
58	2	27/11/2006 17:46:50	http://server.proskin.org/stomp/stomp.aspx	Stomp Questionnaire (reminder)	085219D4
59	2	30/11/2006 12:43:50	http://server.proskin.org/skinprefs/intro.aspx	Skin Preference Questionnaire (reminder)	782C7AD3
60	2	06/12/2006 01:18:51	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	085219D4
61	2	10/12/2006 13:30:00	http://server.proskin.org/ipipneo120/intro120.aspx	IPIPNEO Questionnaire (reminder)	085219D4
(AutoNumber)	0				

Figure 4-9 - Message Centre database table showing the two different types of messaging, User and Questionnaire

Having examined the messaging functionality of the ProSkin platform, the next main application component to be addressed is the skin parser and the user interface skinning system used to provide skins to the Windows Media Player object.

4.11 Changing the User Interface: User Interface Skins

One of the key system requirements (1, from the list of system requirements in Section 3.8) was for an application that featured the ability to change the user interface quickly and easily by participants. In order to meet this requirement two aspects of the platform have been designed to allow this; a Skin Parser component and a ProSkin UML schema.

4.11.1 Skin Parser

The main function of the skin parser is to render user interface skins for the Microsoft Media Player object for the user to interact with. A ProSkin extensible markup language (XML) schema has been defined for use in this research which provides a means to define user interface skins, which the skin parser than renders into ProSkins. User Interface Markup Language (UIML) is an XML language used to define user interfaces (Phanouriou, 2000; Puerta and Eistenstein, 2002; Mayora-Ibarra et al, 2003; Limbourg et al., 2004). One of the benefits of UIML is that it provides the flexibility required to define different ProSkins because it allows the user interface to be

declared in XML as text by element (e.g. background, button, directory window) and then abstracted and applied to construct the actual user interface skin.

4.11.2 ProSkin UIML

All ProSkins are defined in a .proskin file which contains definitions for all user interface skin components. There are three main XML tags used to define a ProSkin: the ProSkinName tag defines the background image to be used and the ProSkinZone tag is then used to define the location, size and image each interface component to appear on the background image and the Stations tag defines the location of the radio station directory. These attributes are summarized below in Figure 4-10 and a shortened example ProSkin is shown in Figure 4-11.

`ProSkinName`	background image name (e.g. blue.jpg)
`ProSkinZone`	interface component name (e.g. play representing the play button)
	location (stated as vectors of X, Y in the format $X_1Y_1X_2Y_2$ e.g. 455,34,63,30)
	state (e.g. normal, mouseover, click),
	location of images per state (e.g. stop_D.gif)

Figure 4-10 - ProSkin XML Tag Definitions

```
<ProSkinCfg>
    <ProSkinName id="1">sp1s4.jpg</ProSkinName>
    <ProSkinZone id="play">
        <Loc>455,34,63,30</Loc>

<Imagetype="normal"default="true">play_D.gif</Image>
        <Image type="over">play_P.gif</Image>
        <Image type="click">play_D.gif</Image>
    </ProSkinZone>
    <ProSkinZone id="stop">
        <Loc>455,80,63,30</Loc>
        <Image type="normal">stop_D.gif</Image>
        <Image type="over">stop_P.gif</Image>
        <Image type="click">stop_D.gif</Image>
    </ProSkinZone>
    <Stations>
        <Loc>58,49,342,149</Loc>
    </Stations>
</ProSkinCfg>
```

Figure 4-11 - Example ProSkin XML File

4.12 Creating ProSkins

The XML schema enables the construction of custom user interface skins that can be created to address the experimental hypotheses, for example colour and shape (as

required by system requirement 6). All interface components are first created in an image manipulation application such as Adobe Photoshop. These include the background image and the interface buttons, which each have up to three states possible (default ("normal"), mouseover ("over") and clicked ("click") that help to provide state feedback to the user. Once the interface graphical components have been created the ProSkin XML definition file can then be constructed.

The first step in the construction of a ProSkin requires the definition of the location of the background image. The interface components are defined as to where they will appear on the background image. In the case of the buttons and status indicator light the ProSkin file also defines the images to be used for the different statuses (e.g. default/mouseover/clicked in the case of the buttons, or in the case of the indicator light green for streaming successful, amber for buffering and red for stream unavailable.

This section has covered all the components of the WebRadio client application. The next section looks at the server side and provides detail of the server architecture.

4.13 ProSkin Server Components

This section looks at the system processes that run on the ProSkin server. These are services that manage participants through the experiment (e.g. allocation of participant to experimental conditions, storage of log file data, timing of ProSkin downloads to clients). Each of these system components are described below in dedicated sections, and illustrated in Figure 4-12:

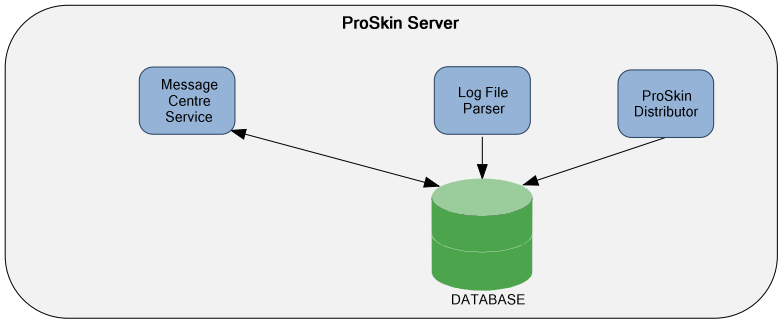


Figure 4-12 - ProSkin Server Architecture

4.13.1 The ProSkin Distributor

The ProSkin Distributor is the system process responsible for controlling the distribution of user interface skins, as specified in system requirement 7. As shall be seen in the experimental chapters 4 to 7, the conditions are defined by the manipulation of experimental variables in the ProSkins, so the ProSkin distributor in practice performs the function of managing participants through the five phases of the experiment. This is because when a participant first installs the WebRadio the client checks with the users table of the database to determine if the user exists. If the user does not exist in the database the ProSkin distributor randomly assigns them to one of the sixteen experimental sequence orders and then transfers the appropriate ProSkins to the client.

In order to prevent participants from only using the radio functionality without contributing questionnaire data, two conditions were programmed into the distributor. These two conditions were a minimum of two radio player uses over a three day time period (use defined for these purposes as opening the web radio) and the completion of the relevant questionnaires. In order to progress from the default install skins ('default' and 'helpskin') phase and to receive a skin pack (from colour, shape and meaning packages), the general and skin preference questionnaires were required to be completed. Additional skin packs were then only limited by time and usage and not by questionnaire completion. At the end of the experiment and having received all skin packages, the user receives all the skin packages (default and helpskins, and skin packages for colour, shape and meaning) and the logging service is terminated to prevent any further data being sent by the client to the server. The process is illustrated in Figure 4-13 overleaf.

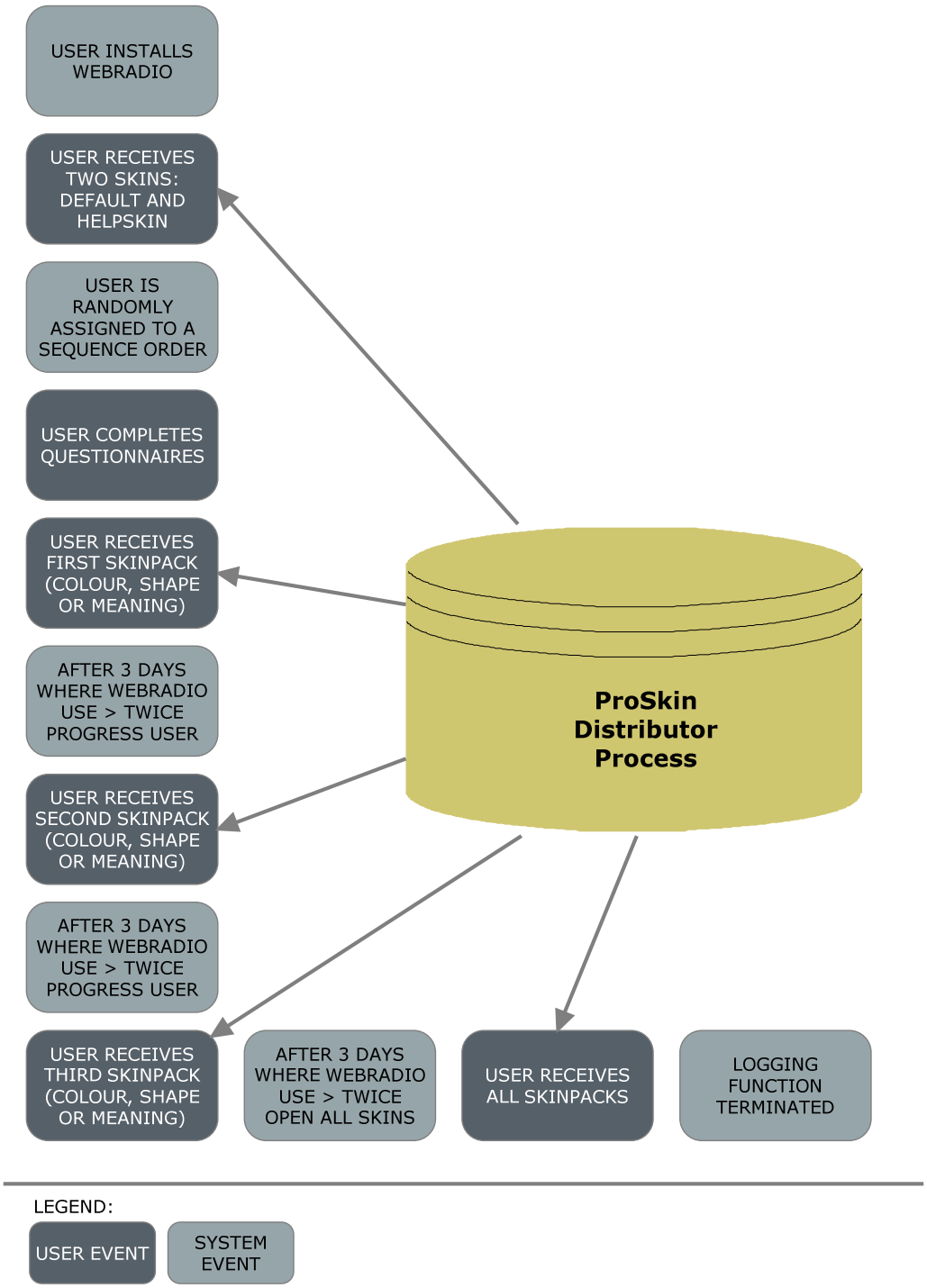


Figure 4-13 - Illustration of the distribution process, with user conditionality shown

4.13.2 Log File Parser

The WebRadio application features a log file recording process to capture and record all user interaction with the application, specifically with regards to the current user interface skin. This meets system requirement 8 for a means to capture user data, both interactive behaviour and questionnaire responses.

When the WebRadio is started a local temporary text file is created for the session and all logs written to that temporary file. At the end of the session, defined as when the user closes the application, the log file is transmitted to the ProSkin server and parsed into the ProSkin database (into the log_header and log_data tables). The application was designed this way to provide the robustness requirement outlined in the system requirements in Section 3.8. By storing the log file locally it provides a faster and more reliable means for initial capture than by transmitting the logged events directly to the database. This is because should the network or Internet connection fail, or be subject to changes in quality (e.g. increased latency, packet drops), the log file data recorded remains intact whilst held locally. If the transfer of the log file is interrupted the client will seek to retransmit the local log files again at a randomly assigned time during the length of the session until successfully sent, or at the initialisation of the application the next time it is run.

4.14 Capturing Interactive Behaviours

There are two main approaches to the capturing and recording of interactive behaviours in log files: event-based logging and time-based logging. Event-based logging involves the capturing of predefined actions when they occur, for example the clicking on a play button (e.g. Liffick and Yohe, 2001; Fine and Brinkman, 2006). Conversely time-based logging involves the capturing of actions at specific predetermined time periods. This research adopts an event-based approach as it appears to be the most appropriate means to capture behaviour over time, as opposed to "snapshot" time sampling.

In order to capture interactive behaviours it is first necessary to define the interaction to be captured. For the purposes of this research, interactive behaviours are defined by the event type and the interface component that the event applies to. For example

when a user clicks on the play button, the click is the event type and the button is the interface component that it applies to. The available interface components and events to be observed are defined in two tables in the ProSkin database, event_type and interface_components. These are summarised in Table 2 and Table 3 below:

Event ID	Event Type	Description
1	Mouse Click	single mouse click
2	Mouse Hover	cursor positioned within the boundaries of the target object but without a mouse click
3	System	used to provide information about the system state in the log file
4	Other	reserved and unused
5	Effort	effort value results stored in log
6	Double Mouse Click	double mouse click (two mouse clicks separated by very short interval, typically <0.25s)
7	Skin Switch	changing the skin via the drop down menu

Table 2 - Summary of event types logged by WebRadio

Interface Component ID	Interface Component	Description
1	Play Button	
2	Stop Button	
3	Setup Button	
4	Hide Button	
5	Volume Up Button	
6	Volume Down Button	
7	Other	reserved and unused
8	Page Up Button	
9	Page Down Button	
10	Open	maximize WebRadio from taskbar minimize
11	Messages	opens the MessageCentre
12	Feedback	opens the anonymous feedback form
13	Exit	terminates WebRadio
14	Systrayicon	the system tray icon
15	Listview	the contents of the radio station directory component
16	ColumnHeaders	The radio station directory column headers

Table 3 - Summary of interface component types logged by WebRadio

Figure 4-14 - Screen capture of the default ProSkin on WebRadio, provided for reference with Table 3 to illustrate the interface components available for interaction by users

The log file process detects when an event is triggered (e.g. a mouse click) and then records the interface component that the event was related to with a timestamp in the log file. The timestamp recorded in the log_data table represents the timing interval between two events and can be used to determine the time between click events. For the purposes of this research, the time between click events has been termed the "interclick" time. An sample extract from the log_data table can be found below in Table 4. It shows the structure of the recorded log data in the database. The columns show a unique session ID per user, a unique user ID, a unique line ID per event, a session timestamp, the event type, the interface component relevant to the event and the interclick time.

log_data							
ld_header_id	ld_user_id	ld_line	ld_timestamp	ld_event_type	ld_int_comp	ld_timing_data	ld_notes
821	389CC1BA	36879	10414.61	2	2	1.031	
724	243884E1	27760	3324.469	2	8	1.516	
724	243884E1	27762	3325.156	2	9	0.656	
842	3CE515B1	39006	157.757	2	4	0.01	
824	68A7DD5A	37176	249.563	2	3	0.031	

Table 4 - Sample extract from the log_data table to illustrate the log file data recorded

4.15 Component Based Messaging

Brinkman, Haakma and Bouwhuis (2001) suggest using the number of messages transferred between software components as a measure of usability. Fewer messages between components indicates a shorter path through the user interface to the task completion and as such can be used to indicate the amount of user effort required to put the system into the desired state. The experimental platform includes program code to identify and count the number of messages transferred between components. By collecting this data two values are generated whenever the user presses the 'Play' button: an ideal and an actual value. The ideal value indicates the smallest number of messages required to get the radio to play the desired station. The actual value indicates the number of messages actually created by the participant to play the desired station. When the two values are the same the user has taken the optimal path. These values are referred to by this research as effort values and are recorded to

determine whether personality type can be identified by effort values so that segments can be formed. A graphical representation of the component architecture is illustrated below in Figure 4-15:

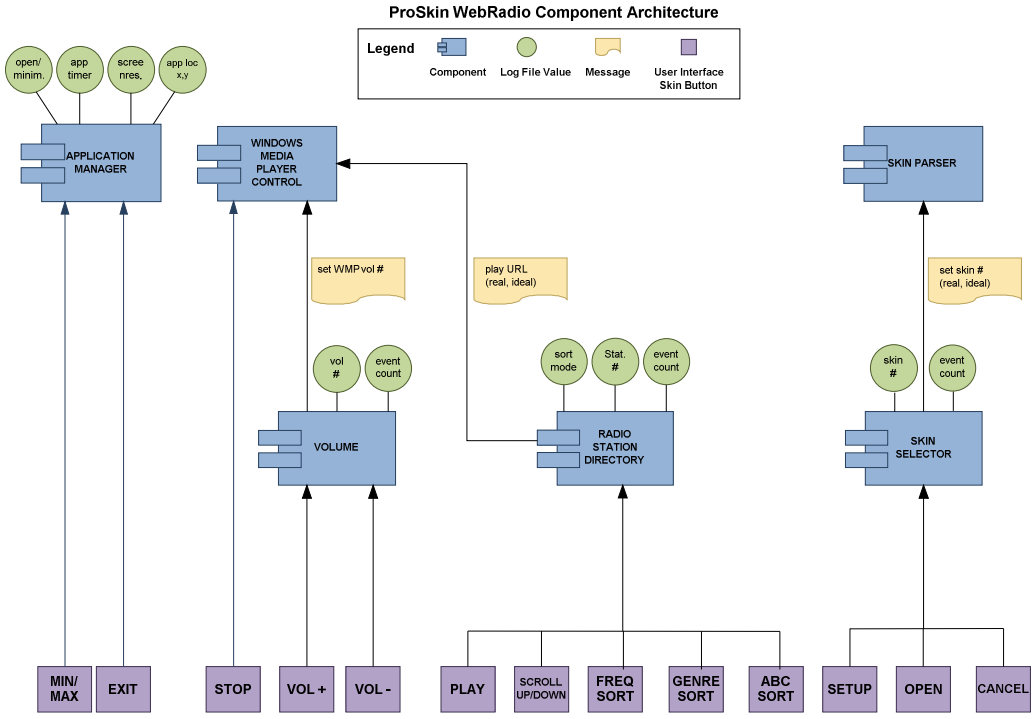


Figure 4-15 - ProSkin Web Radio Component Architecture, depicting components and messages sent between them

This section has shown how interactive behaviours are captured and stored in the experimental platform using log file recording. The next section addresses the ethical issues associated with this research.

4.16 Ethical Considerations

There are a number of ethical issues relevant to this research. The purpose of this section is to define these ethical issues and then to describe how they have been addressed within the design of the experiment and supporting platform.

As with all modern research that involves the use of participants, the experimenter has a duty of care to ensure the safety and wellbeing of participants and as such ethical handling of participants is a requirement of the experimental platform. This research adopts the British Psychological Society's (BPS) Code of Ethics and Conduct (British Psychological Society, 2006) in order to protect participants. The code is based upon

four principles; respect, competence, responsibility and integrity. The adoption and application of these principles within the design of the ProSkin experimental platform will now be discussed.

4.16.1 Privacy

The ProSkin platform handles participant privacy through the anonymous collection of user data. All participants are identified through a non-personally identifiable eight digit alphanumeric string (e.g. 281F80B7). This further helps to ensure participant privacy by protecting participants from being identified by experimenters or in the case of unauthorized access. In order to minimize unauthorized access all participant data is stored in a physically secure data centre and secured at the network level behind an access-limited firewall.

4.16.2 Right of Withdrawal

Participant self determination is addressed by the right of the participant to withdraw from the experiment at any time during the experiment, and there is no pressure applied to the user to either participate or to remain participating against their will. These rights are made explicit both in the installation of the WebRadio application and on the supporting website (www.proskin.org).

4.16.3 Application Specific Recording

The log file process will only record events that occur in relation to the WebRadio application. No other user interactive behaviours in respect of other system applications are logged. Furthermore neither the WebRadio nor the ProSkin server is able to record data from any other of the participant's system applications.

4.16.4 Transparency and Availability of Log File Data

All log files are stored temporarily on the participant's computer. Participants are able to view their recorded data to address any concerns they may have regarding privacy or the nature of data transmitted. Furthermore, the source code for the application is available for inspection on request by participants to ensure that the application is logging the data as claimed.

4.16.5 Consent and Age Checks

All participation in this research is conducted with informed consent from all participants. In addition to this all participants must be over the age of eighteen in order to participate in this research. Both the consent to data collection and age verification checks are performed as part of the WebRadio installation process before the WebRadio is installed. If the participant fails to consent or is not over the age of eighteen then the installation of the WebRadio and participation in the experiment is terminated.

These measures when implemented meet the system requirement 12 in section 3.8 for the ethical handling of participants and participant data. The specific technical details on network security used to protect participant data can be found in the next section.

4.17 Network Architecture

This section describes the design of the network architecture supporting the application described above in section 3.11. One of the benefits of using the Internet to perform research such as this is that the network infrastructure linking participants to the experimenter and experimental platform is already present. All Internet communications are facilitated by the use of the TCP/IP network protocol between client and server, meeting system requirement number 9 on section 3.8.

4.18 Multiple Users

One of the system requirements is an ability to support multiple simultaneous users at any one time (13). This was because when performing an experiment without participants physically located it is difficult to determine how many and when they will participate, so the experimental platform needed to be robust enough to support a realistic number of concurrent users (approximately 10 concurrent users as an estimated maximum). In order to provide this the servers were configured with dual processors, a gigabyte of memory and fast 7200rpm hard drives. The servers were hosted on 100mbps full duplex Internet connections to ensure adequate bandwidth and low network latency. This was necessary so that the experimental platform performed client requests in a timely manner, without causing delays to the users.

There are a number of functions that the experimental platform performs that would have been negatively affected by delayed communications caused by high latency in the connection, hence providing a large 100mbps connection.

These problems include:

- Longer times to start and close WebRadio caused by a delay in communication with the server, causing users to drop out of the experiment.
- Extended download times for new packs of ProSkins, which might cause users to close the application prematurely and jeopardising valid data.
- Extended waiting times for questionnaire items to be saved and presented to the user, resulting in greater levels of incomplete questionnaires.

4.19 Protection of Participant and Experimental Data

The nature of this research means that personal data, albeit non personally identifiable, is stored on the ProSkin server in the database. Ethical guidelines determine that reasonable attempts should be made to protect this data from unauthorised access to protect participant data. In order to provide protection from unauthorised access, firewalls are present on the externally facing network interfaces for both the experimenter network and the ProSkin experimental platform at the hosting facility. The firewall at the experimenter access network is configured to only allow secure terminal services sessions between the experimenter network and the experiment platform network. The firewall at the experimental platform network is configured to allow bidirectional access on the WWW port (80), FTP port (21) and only terminal services connections from the experimenter access network. By doing so these firewall rules provide a reasonable level of protection from unauthorized access and help to ensure the integrity and protection of experimental data, meeting system requirement 12 to protect participant data.

The network architecture is represented in Figure 4-16:

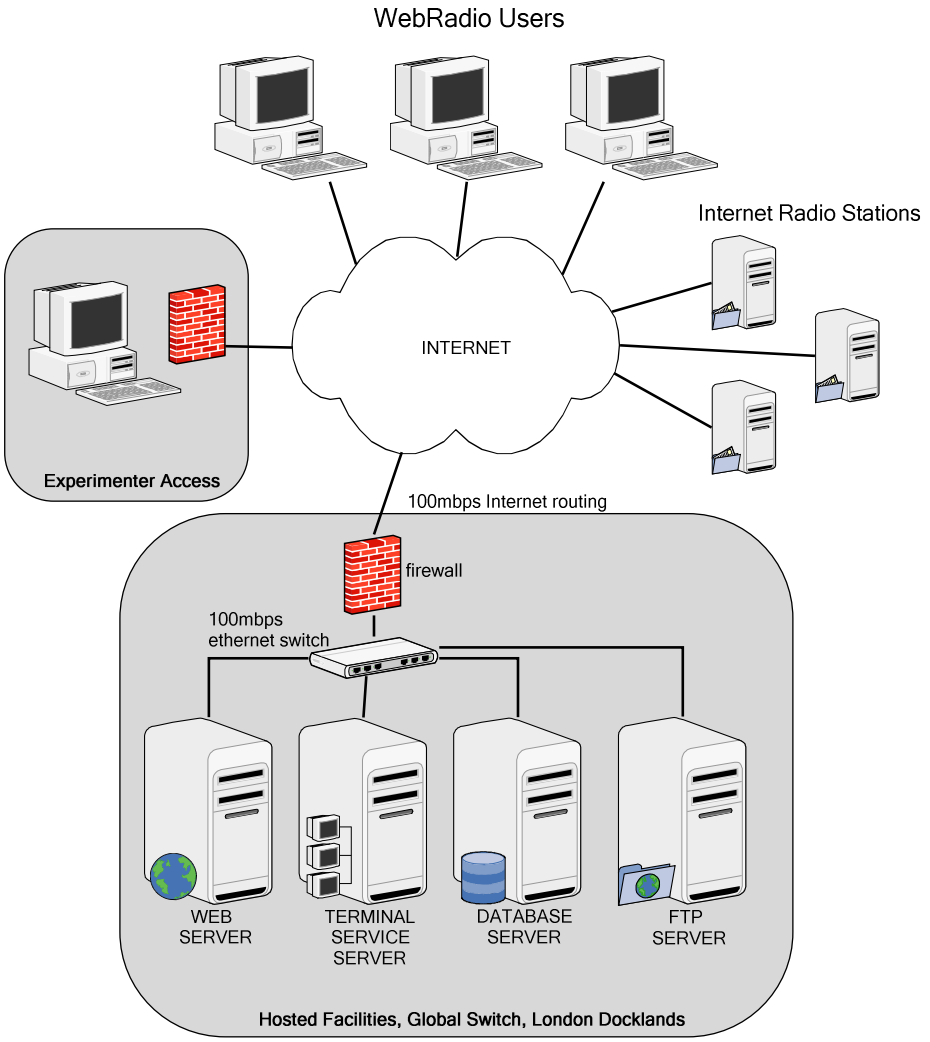


Figure 4-16 - ProSkin Network Architecture

To review, this chapter has described the remote observation approach adopted by this research, and then gone on to describe the design and construction of the client-server architecture required to support this research. As stated at the outset of this chapter, the next four chapters each represent the hypothesis-specific methodological detail and will present hypothesis-specific methodology, results, analysis and discussion, starting out with establishing personality from behaviour, then moving onto potential user interface design factors to be linked to personality traits: colour, shape and meaning.

Chapter 5 – ESTABLISHING PERSONALITY FROM BEHAVIOUR

5.1 Introduction

This chapter covers the first hypothesis of this research, investigating whether interactive behaviours can be used to measure personality traits. As stated in Chapter 2, one of the motivations for this research is to provide a more personally relevant interaction based on personality traits, and that an ability to obtain user personality trait data without the need for the user to complete a psychometric questionnaire prior to the personalisation activity would be of potential benefit to the process of personalisation. The output from this hypothesis will inform the subsequent three hypotheses, because any personality traits that can be identified from recorded log file data can then be supported in the provision of user interface skins appropriate to that personality trait. For example, if extroversion is correlated with clicks per second, and in the subsequent chapter that extroversion is also correlated with the preference for blue user interface skin by extroverts, there are design implications for users with a high click rate who may find a blue skin more personally appealing.

The chapter starts out by briefly summarising the relevant literature support for this hypothesis and then continues by describing the experimental methodology used to investigate H_1. Within the section are details on the construction of behavioural metrics and analysis metrics (e.g. clicks per second, average session time). Following on from the method section is the results section, featuring results and observations of the data analysis. The chapter concludes with sections addressing Limitations and Recommendations for Future Work.

A visual guide to the chapter can be found in Figure 4.1.

H_1: Personality data can be extracted from log files

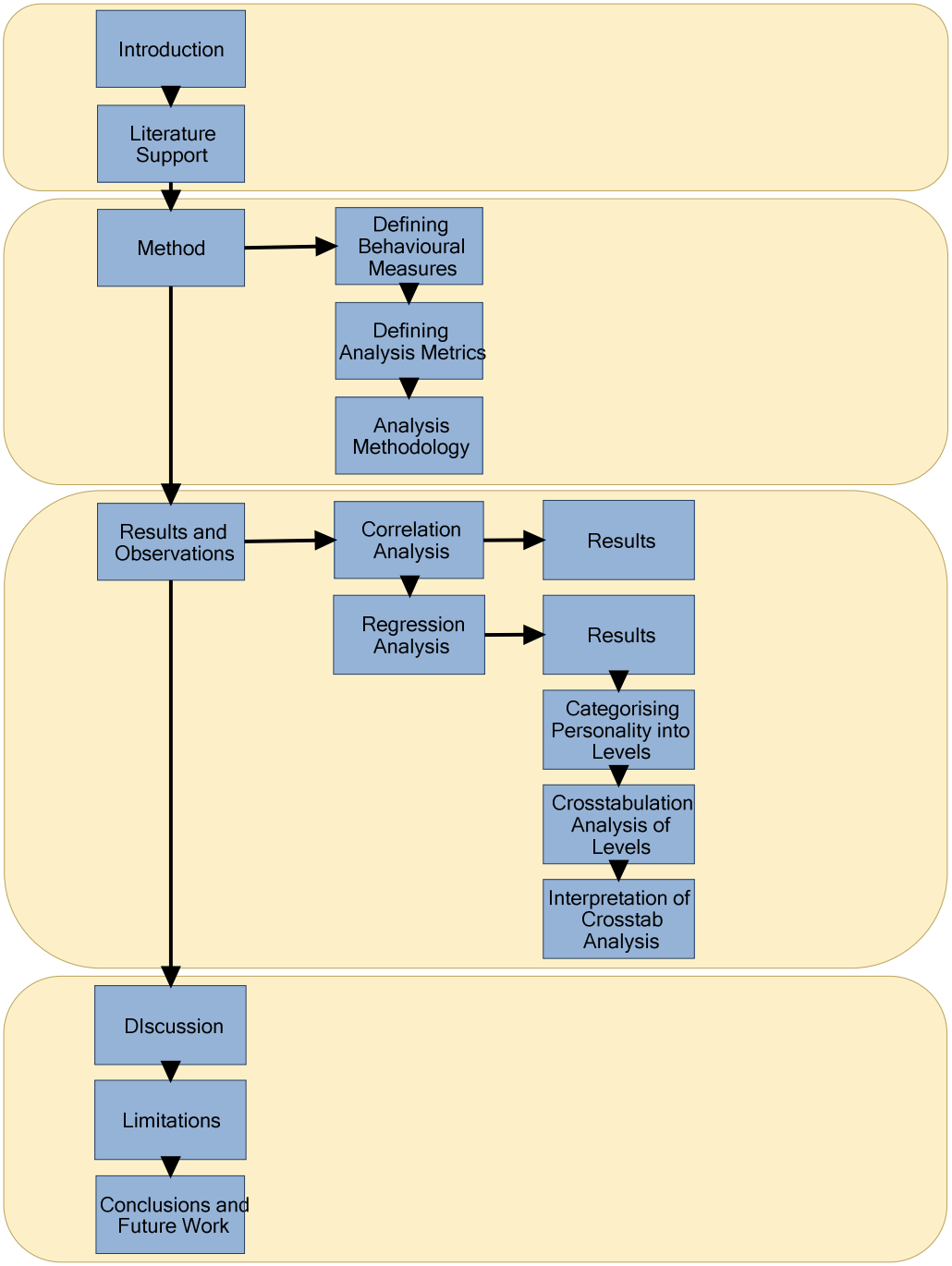


Figure 5-1 - Visual Guide to Chapter 3

5.2 Literature Support

This section briefly reviews evidence from the literature to support the hypothesis, that personality can be determined from interactive behaviour. Further detail regarding these studies and their context within this thesis can be found in Chapter 2.

There are a number of studies that found links between personality and behaviour in general (i.e. not interactive behaviours specifically). In particular links have been found between extroversion and physical movement. Brebner and Cooper (1974) observed extroverts to exhibit higher levels of motor activity with extroverts clicking buttons with a higher frequency when it caused a change in visual stimulation. This would suggest to this research that extroverts could be expected to interact more with the web radio, expressed in a higher click rate (specifically clicks per second, detailed in Section 4.3.1). Whether extroverts click with a higher frequency when it causes an auditory rather than visual change may confirm an additional circumstance under which the higher frequency of clicking by extroverts is observed, as noted by Brebner and Cooper.

Geen (1984) found that extroverts chose more intense noise levels than introverts when selecting a background level to accompany a learning task. This has direct relevance for this research and the experimental platform, because it found links between extroversion and noise levels. This research uses music rather than ambient noise, however it seems logical that a similar effect might be observed with intensity of music levels (expressed as volume). Additionally the task at hand to the user/participant in this research is undetermined as this is externalised research with the participant's task transparent to the experiment, unlike the Geen experiment which was specific to a learning task. Therefore an expectation in investigating the experimental hypothesis is that extroverts might prefer more intense music levels for accompanying computer usage in general rather than learning-specific tasks.

Saati, Salem and Brinkman (2005) found that extroverts tended to interact faster with the user interface than introverts, and replicates reported observations of correlations between extroversion and the speed of human movement (Doucet and Stelmack, 1997). Correlations between interactive speed and conscientiousness as well as neuroticism were observed. These findings suggest that other personality dimensions,

such as Neuroticism and Conscientiousness, in addition to Extroversion might provide additional means to identify personality from interactive behaviour.

Khan et al. (2008) found a number of correlations between personality trait, as measured by the IPIP-NEO inventory, and interactive behaviours when investigating the measurement of Personality from interactive behaviours. They built an application that logged participants keyboard and mouse interaction with any software application on the participant's own computer, and then correlated IPIP-NEO results with behavioural metrics. They defined eleven behavioural metrics, of which three are directly relevant to this research: number of events, average time between events, number of mouse clicks. Correlations were found between these behavioural metrics and a number of personality traits from the IPIP-NEO data, which suggest further investigation and support for H_1.

The next section is the method, which addresses the experimental methodology used to investigate this hypothesis.

5.3 Method

Due to the lack of existing research in this field (log file psychometrics) this study is an explorative study to determine whether personality can be extracted from interactive behaviours recorded in log files (as per H_1). As stated in the previous section there is evidence to suggest that personality traits might be determined from interactive behaviours. It is the intention of this investigation to look for correlations between personality trait and specific interactive behaviours (e.g. volume changes, mouse clicks, session time). This is a similar analysis methodology to both Doucet and Stelmack, and Saati et al.

It is also the intention of this research to establish a behavioural measure of personality from the observation of any potential relationships between personality trait and interactive behaviour. The construct validity of any behavioural measures established will be evaluated by testing predictive validity. Construct validity refers to whether a measure, such as interactive behaviour, measures the psychological construct it was intended to. Predictive validity refers to the ability of a measure to

successfully predict scores on a related measure, in this case the ability to predict personality trait from interactive behaviour.

A total of 143 participants downloaded and installed the ProSkin WebRadio on their own personal computers in their own context, of which 64 provided viable data (55 male, 9 female). They completed a personality inventory (IPIP-NEO) and the results stored on the central ProSkin database. A web radio application was constructed, featuring the ability to record any mouse-based interaction with the webradio by the user and to write the events to a local log file, which was then periodically sent to the central ProSkin server. Where the other hypotheses have included data from three experimental skin packs, this hypothesis only uses the data that relates to the Default skin. The reason for this was because it was the experimental condition that had the highest amount of data, it being one of two skins that all participants received (along with the Helpskin) at installation. The data collection period of the experiment was conducted for a period of three months, during which time users were able to download and use the WebRadio. When the data collection period of the experiment was completed the data was imported to SPSS for analysis.

5.3.1 Defining Behavioural Measures

This section describes the behavioural measures defined from the interactive behaviours expressed through mouse clicks and recorded to a log file by the WebRadio application. The details of the application architecture relating to log file recording can be found in Chapter 3.

The WebRadio application was designed to record any mouse clicks with the default user interface skin. This meant that a number of behavioural measures could be defined from these mouse clicks. Initially four were constructed from the log file data, detailed below in Table 5:

Table 5 - Base set of behavioural measures created

The session time measure was dependent upon the definition of a session, defined for the purposes of this research as the passage of time between two periods of inactivity greater than 30 minutes. The reason why session was defined as such was because as the application output was primarily auditory and therefore a length of time was needed that allowed for non-interaction caused by listening. If the time period was too short it would increase the number of sessions and probably not account for listening. However, if the time period was too long it would not only decrease the number of sessions, but it would also not account for users leaving the PC that the web radio was installed on, with the web radio still playing. As a result of this a period of 30 minutes of inactivity seemed like a reasonable time period that handled listening time and also as a time out for user absence as the experiment required interaction data and not non-interaction during listening. It also provides a more conservative measure by attempting to control for non-interaction time.

In addition to the original four measures, an additional three measures were constructed. The first was clicks per second, defined as total clicks divided by total session time to give this additional measure of rate.

The WebRadio application also captured effort values relating to the radio station directory. The idea of component based usability comes from Brinkman, Haakma and Bouwhuis (2007), who suggest a component-based approach to software application usability evaluation. The concept of effort values to change the state of the machine via software constructs, defined as components, is adopted in this research as a behavioural metric. The capturing of effort values relating to station change meant that whenever the user interacted with the radio station directory to select a station to play, whenever the play button was pressed two effort values were generated by the application and recorded to the log file – an ideal and an actual effort value. The ideal effort value represented the shortest path to the users intended radio station in the directory and the actual effort value represented the actual length of the path to the intended radio station. The difference between the two was caused by the multiple navigational paths available to the user to be able to sort the radio station directory to be able to search through the available radio stations. For example, the user could sort the radio station directory by name, genre, language, number of plays and last played, as well as scrolling the entire list vertically using scroll buttons. This

provided sufficient diversity in navigational paths for the user to generate the differential values, ideal and actual effort. From these two values, an Extra Effort measure was created. This was defined as actual effort minus ideal effort, representing the additional extra effort used by the user to reach the finally selected radio station. Two values were created from extra effort, a mean measure and a sum total measure, per session for each station change. These metrics were chosen because they represented a broad range of interactive behaviours, including physical action (clicks), time measures (session time, clicks per second), feature changes (volume, station) and user effort (extra effort).

Table 6 - Final set of seven behavioural measures

5.3.2 Developing Analysis Metrics for Defined Behavioural Measures

Having defined a set of seven behavioural measures an additional set of metrics were defined for the data analysis. Each of the metrics was defined so that they could be applied to the seven behavioural measures above. These metrics were chosen because they in part represent descriptive statistics (count, average, standard deviation, total) and also they provide the data required to describe interaction in numerical and graphical terms (intercept, slope, r2). This is useful because it allows the analysis of the rate of interactive behaviours.

The concept of sessions as a unit for analysis allowed a linear function to be applied over all sessions per user for each of the behavioural measures by regression. By applying this linear function for to all behavioural measures it allowed session data to be aggregated for analysis and metrics to be defined (e.g. slope, intercept, r^2). An example is shown below in Figure 5-2 to illustrate the linear regression over sessions:

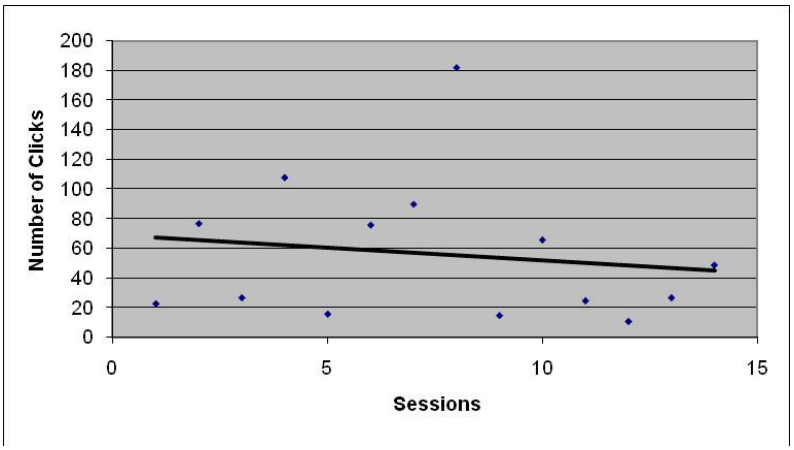


Figure 5-2 - Example scatter plot for Number of Clicks over Sessions for a single participant, showing linear function applied (r=-0.15, slope=-1.70, intercept=69.40)

The list of metrics are listed and described in Table 7:

Metric	Description
Count	A count of the number of sessions
Average	The mean of the behavioural measure across all sessions
Standard Deviation	The standard deviation of the behavioural measure mean
Total	The total number of behavioural measure events across all sessions
Intercept	The intercept of the linear function fitted to the data across all sessions
Slope	The slope of the linear function fitted to the data across all sessions
r^2	The amount of variance accounted for by the linear function

Table 7 - Summary and description of analysis metrics to be applied to behavioural measures

The values for count represent the count of the number of events per session. As a result of this the count values for total clicks, session time, clicks per second, station change and volume change are all the same. They may be referred to collectively as event count (Ev_Count). However, the count for extra effort (mean and sum) represent a different count to the other previously mentioned because they are generated when the radio station is changed. Not all users changed the radio station sub-optimally, or at all, which meant that no effort value was generated. Only those users with effort values therefore generated an effort count.

These analysis metrics, when applied to each of the behavioural measures, produced the following analysis metrics for this hypothesis, listed in Table 8. The details of the

analysis methodology are provided in the section following this one.

Total Clicks	Session Time	Clicks Per Second	Station Change	Volume Change	Extra Effort Mean	Extra Effort Sum
Ev_Count	Ev_Count	Ev_Count	Ev_Count	Ev_Count	EEM_Count	EES_Count
TC_Average	ST_Average	CPS_Average	SC_Average	VC_Average	EEM_Average	EES_Average
TC_SD	ST_SD	CPS_SD	SC_SD	VC_SD	EEM_SD	EES_SD
TC_Total	ST_Total	CPS_Total	SC_Total	VC_Total	EEM_Total	EES_Total
TC_Intercept	ST_Intercept	CPS_Intercept	SC_Intercept	VC_Intercept	EEM_Intercept	EES_Intercept
TC_Slope	ST_Slope	CPS_Slope	SC_Slope	VC_Slope	EEM_Slope	EES_Slope
TC_r2	ST_r2	CPS_r2	SC_r2	VC_r2	EEM_r2	EES_r2

Table 8 - Summary of all analysis metrics used to investigate H_1

5.3.3 Analysis Methodology

The analysis methodology for this hypothesis was to perform correlation analysis (Pearson's) between personality trait data collected from the IPIP-NEO questionnaire in the WebRadio with each of the behavioural metrics defined in Table 8, captured and collated by the WebRadio client and ProSkin server respectively. This would potentially provide a set of significant correlations indicating relationships between certain personality traits and behavioural measures.

For any personality trait correlations that were significant for more than one analysis metric, a regression analysis was performed using a stepwise method using each significant analysis metric as the independent variable and the personality trait as the dependent variable. For example, if Conscientiousness was found to be significantly correlated ($p=<0.05$) with TC_Total and CPS_Average, these two behavioural metrics would have been entered into the stepwise model, with the best model selected (as defined as the highest r^2 value). The r^2 value derived from the regression analysis would provide the amount of variance in personality trait data accounted for by the behaviour data.

The results from the regressions and the subsequently constructed predictive models were then evaluated for construct validity to establish how well the behavioural metrics measured what they were designed to measure, Personality. In particular predictive validity was tested using crosstab analysis to evaluate the predictive success of the regressions models.

5.4 Results and Observations

The results of the correlation analysis provide the data required to construct a predictive model and for the purposes of this research are considered to be an interim stage of the analysis. They do however offer a number of early insights into any potential relationships between personality trait and interactive behaviours. Interpretation of these correlations when represented in tables of data such as these is difficult because there are a number of traits where there are multiple correlations with behavioural metrics. The correlation results have been translated into English statements to provide another perspective to view the results and appear beneath each of the summary tables for each trait.

In interpreting the results of the correlations it is difficult to easily understand certain personality traits as they relate to interactive behaviours, for example friendliness. As a result of this only the correlations that appear to have face validity are interpreted. Face validity is a form of construct validity that describes whether the measure appears to measure the intended construct (Anastasi, 1988). Face validity is inherently subjective, but whilst this subjective interpretation has its inherent bias, within the context of exploratory research into an under researched area it may provide insights and direction for future work. These interpretations are drawn informally for any tentative early understanding and are not intended to directly test the experimental hypothesis (H_1).

There was a large difference in the gender distribution of participants for this experiment (55 males, 9 females). As a result of this the data for the females was not included in the analysis and any conclusions drawn limited to males.

5.4.1 Correlation Analysis

5.4.1.1 *Extraversion*

Behavioural Metric	Personality Trait									
	Extraversion	n	Friendliness	n	Gregariousness	n	Assertiveness	n	Cheerfulness	n
TC_SD	-0.478**	30	-0.372*	30					-0.384*	30
ST_r2			0.380*	30						
SC_SD	-0.463**	30			-0.374*	30	-0.472**	30	-0.366*	30
EES_Average							-0.321*	44		
EES_SD	-0.426*	22								
EES_Intercept							-0.307*	44		

* indicates significant at <0.05 (two tailed), ** indicates significant at <0.01(two tailed)

no significant correlations for the Excitement Seeking or Activity Level facets

Table 9 - Summary of significant correlations for Extraversion traits

Personality Trait	Relationship between Personality Trait and Behavioural Metric
Extraversion	more consistent in their total number of clicks, station changes and total extra effort
Friendliness	more consistent in their total number of clicks, station changes and total extra effort
	the amount of variance that can be explained by session time increases with friendliness
Gregariousness	more consistent in their total number of station changes
Assertiveness	more consistent in their total number of station changes
	as assertiveness increases, average total extra effort and the intercept of total extra effort decreases
Cheerfulness	more consistent in their total number of clicks and station changes

Table 10 - Correlation analysis results for Extraversion expressed as English statements

The main extraversion trait showed a predominance of consistency in clicks, station changes and total extra effort, suggesting that extroverts are more consistent in their interactive behaviour than introverts. There is no indication from the data as to the direction of this consistency, for example whether they consistently make more station changes. Furthermore, one might expect highly extrovert individuals to exhibit higher clicks per second, but this was not supported by the data. These results are somewhat disappointing in that there are a relatively few significant correlations compared to the other traits and also because the Extraversion factor had the most amount of literature supporting expression of this personality trait in behaviour.

These results do not provide any evidence to support the findings of Saati, Salem and Brinkman (2005) who found extroversion negatively correlated with speed (i.e. faster) and positively correlated with session events when interacting with a software based CD player. Furthermore, they did not replicate the findings of Khan et al. (2008). None of the correlations observed by Khan et al. were observed from this data. There are a number of reasons as to why this occurred. The Khan methodology featured

multiple application logging, whereas the Proskin application featured a single application (WebRadio) for behavioural analysis. The WebRadio interface was designed to be a mouse only interface without requiring any keyboard input to operate the radio functionality. Other applications, such as word processors, development environments, e-mail clients and web browsers, frequently require keyboard input. An explanation for the difference in the findings of this experiment and those of Khan et al. is differing interactive behaviours as a function of keyboard and mouse combinations. Perhaps the dual input requirement and the changes between these two modes of input accounts for the lack of supporting evidence found for Khan et. al. in this experiment. Another reason for the differences is partially explained by the differences in definition of time between mouse clicks between the Khan research and this experiment. Khan et al. applied a filter to their mouse click data to remove any entries within two seconds of each other, to remove double clicks and user mistakes. This meant that effectively their events per session metric were limited to a minimum of two seconds, whereas this experiment did not apply such a filter. This difference in the available range of mouse clicks per session may have contributed to the lack of support found for Khan et al.

The face validity of these results suggest that there is a possible consistency in behaviour of highly extrovert individuals in comparison with low extroversion (i.e. introvert) individuals.

5.4.1.2 *Agreeableness*

Personality Trait (Agreeableness dimension)								
Behavioural Metric	Agreeableness	n	Trust	n	Altruism	n	Modesty	n
Ev_Count					0.283*	49		
TC_r2	-0.448*	30						
ST_Average							-0.323*	49
ST_SD							-0.382*	30
ST_Total							-0.315*	49
ST_Intercept							-0.332*	49
ST_Slope					0.326*	49		
CPS_SD					-0.403*	30		
SC_SD			-0.402*	30				
VC_Total	0.299*	49						
VC_r2	-0.611**	25			-0.549**	30		
EEM_Count					0.283*	49		
EEM_r2					-0.499*	21		
EES_Count					0.283*	49		
EES_r2					-0.536*	20		

* indicates significant at <0.05 (two tailed), ** indicates significant at <0.01(two tailed)
no significant correlations for the Cooperation, Sympathy or Morality facets

Table 11 - Summary of significant correlations for Agreeableness traits

Personality Trait	Relationship between Personality Trait and Behavioural Metric
Agreeableness	the amount of variance that can be explained by total clicks and volume changes decreases as agreeableness increases
	as agreeableness increases, the amount of volume changes increase
Trust	more consistent in their total number of station changes
Altruism	as altruism increases the total number of session events increases (count)
	the slope of session time increases as altruism increases
	more consistent in their total number of clicks per second
	average extra effort and total extra effort decreases as altruism increases
	the amount of variance that can be explained by volume changes
Modesty	as modesty increases, average session time, total session time and the intercept of session time decreases
	more consistent in their total session time

Table 12 - Correlation analysis results for Agreeableness expressed as English statements

Within the agreeableness factor, altruism showed a tendency for highly altruistic people to interact more with the web radio, having a significant correlation with session events. This is possibly explained by the scenario in which altruistic people wanted to help the investigator and therefore consciously using the web radio more in order to acquiesce to any perceived needs. Some evidence to support this is also found in the positive correlation between the slope of session time and altruism. However, no explicit evidence in the form of positive correlations with session time or extra effort was found. The modesty factor showed a tendency for very modest people to interact less with the web radio. They had shorter average session times and shorter total session times in particular and also displayed a significantly negative correlation with standard deviation, implying more consistency in these session times.

These findings make sense given that modest people are characteristically limited in their behaviours, expressed in consistently limited session times.

5.4.1.3 *Conscientiousness*

Beh. Metric	Personality Trait (Conscientiousness dimension)													
	Consc'ness	n	Self Efficacy	n	Orderliness	n	Dutifulness	n	Ach. Striving	n	Self Discipline	n	Cautiousness	n
TC_Count	0.358*	49	0.301*	49										
TC_Total	0.346*	49	0.329*	49									0.324*	49
TC_r2	-0.467**	30	-0.427*	30					-0.440*	30				
ST_Count	0.358*	49	0.301*	49										
ST_r2	-0.471**	30	-0.443*	30									-0.371*	30
CPS_Count	0.358*	49	0.301*	49										
CPS_r2	-0.483**	30	-0.418*	30	-0.362*	30			-0.573**	25				
SC_Count	0.358*	49	0.301*	49										
SC_Total	0.285*	49	0.297*	49									0.301*	49
VC_Count	0.358*	49	0.301*	49										
VC_r2	-0.704**	25	-0.578**	25							-0.550**	25		
EEM_Count	0.381*	44	0.414**	44										
EEM_Total	0.384*	44					0.304*	44						
EEM_Slope	-0.332*	44					-0.376*	44						
EEM_r2			-0.571**	21										
EES_Count	0.381*	44	0.414**	44										
EES_Total	0.382*	44	0.331*	44									0.345*	44
EES_Intercept													0.322*	44
EES_r2			-0.611**	20					-0.469*	20				
* indicates significant at <0.05 (tw o tailed), ** indicates significant at <0.01(tw o tailed)														

Table 13 - Summary of significant correlations for conscientiousness traits

Personality Trait	Relationship between Personality Trait and Behavioural Metric
Conscientiousness	as conscientiousness increases the total number of session events, total number of clicks increases (count)
	as conscientiousness increases the total amount of station changes, average and total extra effort increases
	the amount of variance that can be explained by total clicks, session time, clicks per second and volume changes decreases as conscientiousn
	the slope of average extra effort decreases as conscientiousness increases
Self Efficacy	as self efficacy increases the total number of session events,total number of clicks, total station changes, total extra effort increases (count)
	the amount of variance that can be explained by total clicks, session time, clicks per second, volume changes,average extra effort…
	...total extra effort decreases as self efficacy increases
Orderliness	the amount of variance that can be explained by clicks per second decreases as orderliness increases
Dutifulness	as dutifulness increases the amount of total average extra effort increases
	the slope of average extra effort decreases as dutifulness increases
Achievement Striving	the amount of variance that can be explained by total clicks, clicks per second, total extra effort decreases as achievement striving increases
Cautiousness	as cautiousness increases the total number of clicks, total number of station changes, total extra effort and the intercept of total extra effort incr
	the amount of variance that can be explained by session time decreases as cautiousness increases

Table 14 - Correlation analysis results for conscientiousness expressed as English statements

Conscientiousness, as a main factor, was positively correlated with a number of behavioural metrics. They suggest a tendency for conscientious individuals to interact more with the web radio, in the form of total number of session events, more total clicks, more station changes and greater total extra effort. The slope of average extra effort also decreases for highly conscientious people, which suggests that the attention to detail that often characterises the conscientious trait might have resulted in achieving the required state more efficiently over time and hence less extra effort required.

Self efficacy, a facet of conscientiousness, like conscientiousness above was correlated with a number of behavioural metrics. The majority of the correlations for both self efficacy and conscientiousness are the same and this is possibly because of a similarity in the underlying construct that these two traits represent – especially in the context of interactive behaviour towards a web radio. As with highly conscientious individuals, highly self efficacious individuals showed a tendency for more session events, more station changes, more clicks and more total extra effort events. It might be that their efficacy is demonstrated in increases interaction with the web radio to show "better" participation in the experiment.

The dutifulness factor showed a tendency for total average effort to increase for highly dutiful people. Like altruism previously this might be a confound due to dutiful people feeling a compulsion to assist the investigator from this sense of duty. This is exhibited in average extra effort, showing perhaps a willingness to explore the web radio more. However, the slope of average extra effort decreases with dutifulness, suggesting that perhaps this sense of duty is transitory and wanes over time.

Cautiousness showed increases in total number of clicks, station changes, total extra effort and the intercept of total extra effort as a function of cautiousness. These findings are all contrary to expectation, which is that highly cautious individuals will interact less with the web radio. It is difficult to explain this finding and further research is required to understand why. It might be that highly cautious people try a lot of different things first experimentally before settling down into a preferred behavioural pattern, for example trying out a lot of different radio stations and spending more additional effort to explore the radio, which is supported by the data.

5.4.1.4 *Neuroticism*

Behavioural Metric	Personality trait (Neuroticism dimension)									
	Neuroticism	n	Anger	n	Self Consciousness	n	Immoderation	n	Vulnerability	n
TC_Count										
TC_Average			-0.327*	49						
TC_SD			-0.421*	30	0.495*	30				
TC_Intercept			-0.286*	49						
TC_r2			-0.370*	30						
ST_Average									-0.283*	49
ST_r2			-0.409*	30						
CPS_Average			-0.296*	49						
CPS_SD			-0.368*	30						
CPS_Total			-0.309*	49						
CPS_Intercept			-0.302*	49						
SC_SD					0.390*	30	-0.462*	30		
SC_Intercept							-0.317*	49		
VC_Average							0.295*	49		
VC_r2	-0.437*	25	-0.512**	25	-0.418*	25				
EEM_Average									-0.337*	44
EEM_SD			-0.452*	22						
EEM_Intercept									-0.321*	44

* indicates significant at <0.05 (two tailed), ** indicates significant at <0.01(two tailed)

no significant correlations for the Depression facet

Table 15 - Summary of significant correlations for Neuroticism traits

Personality Trait	Relationship between Personality Trait and Behavioural Metric
Neuroticism	the amount of variance that can be explained by volume changes decreases as neuroticism increases
Anger	as anger increases, average total clicks, intercept of total clicks, average clicks per second, total clicks per second (cont.)
	(cont.) the intercept of clicks per second decreases
	more consistent in their total number of clicks, clicks per second, average extra effort
	the amount of variance that can be explained by total clicks, session time and volume change decreases as anger increases
Self Consciousness	more consistent in their total number of clicks, station changes
	the amount of variance that can be explained by volume changes decreases as self consciousness increases
Immoderation	more consistent in their station changes
	the intercept of station changes decreases as immoderation increases
	as immoderation increases the average number of volume changes increases
Vulnerability	as vulnerability increases, average session time, average extra effort and the intercept of average extra effort decreases

Table 16 - Correlation analysis results for Neuroticism expressed as English statements

Anger, a facet of neuroticism, showed that as anger increases interaction with the web radio decreased, particularly in the form of clicks and clicks per second (as opposed to session time, station change, volume change etc). This was represented by negative correlations with average total clicks, the low number of total clicks in the first session, average clicks per second, total clicks per second and the low number of clicks per second in the first session. One interpretation of these findings is that highly angry people are more engaged by music, possibly by an emotional connection, and as such requires lower interaction with the web radio. This also might be considered an experimental confound because listening to the radio whilst not interacting with it is still strictly an interaction, but yet is not explicitly measured by this research. Also of note is the exclusively negative array of correlations for

anger, which suggests a trend in lower interactivity with the web radio in general as a function of anger.

Self consciousness showed a tendency for highly self conscious individuals to be more consistent in their interaction with the web radio in respect of total number of clicks and station changes. This might be interpreted as self conscious individuals being more stable and consistent in their interaction with the web radio because of the awareness that all interactions with the web radio application were logged.

Immoderation showed that as immoderation increased so did the average number of volume changes. It might be that highly immoderate people adjust the volume up and down more. This behaviour would be characteristic of a typically immoderate individual that exhibited excessive behaviour, such as preferring loud music volume. However, other behavioural metrics that also could indicate excessive interactive behavioural tendencies, such as session time and session events, did not reveal any significant correlations.

Vulnerability showed a tendency for highly vulnerable individuals to spend less time on average interacting with the web radio. As for self consciousness, the fact that the web radio application logged interaction might have been cause for concern to highly vulnerable individuals as a perceived threat to privacy.

5.4.1.5 *Openness To New Experience*

Behavioural Metric	Personality trait (Openness to New Experience dimension)									
	Imagination	n	Artistic Interests	n	Emotionality	n	Intellect	n	Liberalism	n
TC_SD			-0.419*	30						
TC_Total			-0.307*	49						
TC_Intercept			-0.317*	49						
TC_Slope			0.314*	49						
ST_Average	-0.311*	49					0.311*	49		
ST_SD									-0.368*	30
ST_Total							0.289*	49	-0.304*	49
ST_Intercept	-0.322**	49					0.302*	49	-0.282*	49
ST_Slope					0.318*	49				
SC_SD					-0.377*	30				
SC_Slope			0.312*	49						
EES_Slope			0.326*	44						

* indicates significant at <0.05 (two tailed), ** indicates significant at <0.01(two tailed)

no significant correlations for the main Openness To New Experience factor, or the Adventurousness facet

Table 17 - Summary of significant correlations for openness to new experience traits

Personality Trait	Relationship between Personality Trait and Behavioural Metric
Artistic Interests	more consistent in their total number of clicks
	as artistic interests increases, the total number of clicks and the intercept of total clicks decreases
	the slope of total clicks, station change, and extra effort all increases as artistic interests increases
Emotionality	the slope of session time increases as a emotionality increases
	more consistent in their total number of station changes
Intellect	as intellect increases, average session time,total session and intercept of session time increases
Liberalism	more consistent in their total session time
	as liberalism increases, total session time decreases
	the intercept of session time decreases as liberalism increases

Table 18 - Correlation analysis results for openness to new experience expressed as English statements

Liberalism, a facet of openness to new experience, showed a tendency for a consistency in total session time and increased session times. This finding appears to have face validity because by definition the liberal individual is likely to express greater interaction and according to this data expresses that liberalism in the form of increased total session time.

5.4.2 Methodology for Construction of Predictive Models

The correlation analysis determined which personality traits were most likely related to certain of the behavioural metrics. These metrics would be then used in a regression model to determine the amount of variance from each contributing personality trait that could be explained by the behavioural data.

The personality traits were each entered into a linear regression analysis using a stepwise method, with the dependent variable as the personality trait and the independent variables as the behavioural metrics listed in Table 8. The output from these regressions provide an indication of how well behaviour could explain variance for each personality trait (where possible) with further detail about how much of the variance is accounted for by each trait. This provides an indicator as to how accurately the model will perform as a predictor of personality trait from behavioural metric/s. Furthermore, a model with multiple predictors would suggest that a combination of behavioural measures could increase the accuracy of the personality measurement.

5.4.2.1 *Results*

The results of the regression analysis can be found in Table 19 below:

Personality Trait	Regression Results								
	r	r2	adj r2	Std. Error	F	sig	df	component1	component2
Extraversion	0.545	0.297	0.262	20.31	8.453	0.009	1,20	TC_SD	
Friendliness	0.533	0.284	0.231	22.44	5.352	0.043	2,27	ST_r2	TC_SD
Gregariousness	0.374	0.140	0.109	27.41	4.557	0.011	1,28	SC_SD	
Assertiveness	0.563	0.317	0.292	20.49	12.557	0.001	1,47	TC_SD	
Activity Level	none								
Excitement Seeking	none								
Cheerfulness	0.384	0.147	0.117	22.98	4.841	0.036	1,28	TC_SD	
Agreeableness	0.611	0.373	0.346	17.46	13.680	0.001	1,47	VC_r2	
Trust	0.402	0.161	0.131	22.48	5.388	0.028	1,28	SC_SD	
Morality	none								
Altruism	0.564	0.319	0.288	21.40	10.282	0.004	1,22	VC_r2	
Cooperation	none								
Modesty	0.382	0.146	0.115	25.97	4.772	0.037	1,28	ST_SD	
Sympathy	none								
Conscientiousness	0.698	0.487	0.463	17.03	20.846	0.000	1,22	VC_r2	
Self Efficacy	0.637	0.406	0.366	23.07	10.239	0.006	1,15	EES_r2	
Orderliness	0.362	0.131	0.100	29.11	4.228	0.049	1,28	CPS_r2	
Dutifulness	0.376	0.141	0.121	24.45	6.918	0.012	1,42	EEM_Slope	
Achievement Striving	0.539	0.290	0.251	21.24	7.367	0.014	1,18	TC_r2	
Self Discipline	0.311	0.096	0.064	23.67	2.988	0.095	1,28	CPS_r2	
Cautiousness	0.382	0.146	0.115	23.84	4.623	0.041	1,27	EES_Intercept	
Neuroticism	0.437	0.191	0.156	24.39	5.431	0.029	1,23	VC_r2	
Anxiety	none								
Anger	0.651	0.423	0.390	22.27	12.488	0.003	1,17	TC_Average	
Depression	none								
Self Consciousness	0.683	0.466	0.418	22.73	5.455	0.001	2,22	TC_SD	VC_r2
Immoderation	0.499	0.249	0.216	25.84	7.608	0.001	2,46	SC_Intercept	VC_Average
Vulnerability	0.337	0.113	0.092	24.18	5.368	0.025	1,42	EEM_Average	
Openness	none								
Imagination	0.322	0.104	0.085	30.08	5.433	0.024	1,47	ST_Intercept	
Artistic Interests	0.569	0.324	0.299	23.92	12.946	0.001	1,27	TC_Intercept	
Emotionality	0.377	0.142	0.111	26.75	4.635	0.040	1,28	SC_SD	
Adventurousness	none								
Intellect	0.311	0.097	0.077	24.27	5.032	0.030	1,47	ST_Average	
Liberalism	0.368	0.135	0.104	29.32	4.373	0.046	1,28	ST_SD	
TC=total clicks									
SC=station change									
ST=session time									
VC=volume change									
EEM=extra effort mean									
EES=extra effort sum									
CPS=clicks per second									

Table 19 - Summary of regression analysis

The results of the regressions show that the best model to predict personality trait from behavioural metrics is Conscientiousness with an r^2 of approximately 48%. This means that at best the maximum variance in a trait that could be explained by behavioural data was approx 48%, with four traits with r^2 values greater than 40%. These scores can be considered low when trying to successfully predict personality from behaviour because at least 51% of the variance remains unaccounted for in the highest r^2 scoring trait (Conscientiousness). Furthermore, the standard error for all of these regression models is approximately 24 points (mean 23.74, n=26). The IPIP-NEO output for all five main personality factors is on a 100-point scale, and so a standard error of 24 points indicates an average deviation of 24 points between prediction and actual IPIP-NEO. This error is too high for such regression models to be accurate enough to be able to produce useful predictions of personality trait on a 100 point scale measurement instrument.

However, one must also consider the practical implications of any output from a measurement instrument as generated by this work. Ideally a highly accurate measurement would be able to be made from either a single or multiple behavioural measure (e.g. $r^2 \geq 0.8$). In such a circumstance the input to a designer to be able to implement a user interface skin to accommodate a predicted personality trait might be on a 100 point scale, but the designer would not be reasonably or practically expected to produce 100 user interface skins, one for each point on that particular trait scale. In order to accommodate the differences between different levels of a trait, a more simplified approach is most likely: to produce skins for low, medium and high levels of any particular trait. As a result of this, the prediction model does not need to be as accurate because it does not need to exactly replicate the accuracy of IPIP-NEO to be of use as a methodology for deducing trait from behaviour. The intention of this work is to identify traits from behaviour so that appropriate user interface skin can be produced, rather than to produce a behavioural measure that is highly correlated with IPIP-NEO as a means for direct replacement. In order to test the predictive ability of the regression models, the results were tested using categorised personality trait levels (low, medium and high) to reflect the context for usage. The next section will address this in further detail.

5.4.2.2 *Categorising Personality Trait Into Levels*

As stated in the previous section, the standard error produced by the regressions was approximately 24 points and as such prevented a highly accurate prediction of personality trait from behaviour. The alternate approach was to recode the personality trait scores into low (0-33.333 points), medium (33.334-66.666 points) and high (66.667-100 points) categories. These category breaks were selected according to the breaks used by the IPIP-NEO inventory (Buchanan et al., 2005). By recoding into categories of high, medium and low in this manner it facilitates identification of category of personality trait (e.g. low extrovert) and better accommodates the need of designers and anyone who will be producing personalised user interface skins from trait data.

5.4.2.3 *Crosstabulation Analysis of Personality Trait Categories*

In order to test the predictive ability of the regression models against the categorised personality traits, the unstandardised predicted personality trait score was calculated. These predicted trait scores were then recoded and categorised in the same way as the original trait scores were, as described in the previous section. This enabled a crosstabulation analysis to compare the original trait scores with the predicted trait scores (based on the regression models) under low, medium and high trait score conditions. The crosstabulation analysis is used to examine the predictive ability of the regression models under the new categorised trait structure.

An example trait, self consciousness, was selected for the crosstabulation to illustrate the predictive ability of the regressions because it showed a high r^2 and used two components in the model. The results are shown in Table 20:

Count

		Predicted Self Consciousness			
		low	medium	high	Total
Original Self Consciousness	low	5	4	0	9
	medium	0	4	1	5
	high	1	5	5	11
Total		6	13	6	25

Table 20 - Crosstabulation for Original Self Consciousness * Predicated Self Consciousness

5.4.2.4 *Interpretation of Crosstabulation Analysis of Categorical Personality Trait Data for Self Consciousness*

The results for the self consciousness data indicate that out of a total of 25 predictions, 14 were accurate (56%). This hit rate is calculated by the sum of the correct values divided by the total number of predictions, so for self consciousness the hit rate of 56% was calculated from 5+4+5=14/25=56%. Error rates can be considered on two different levels, depending upon the severity of impact. Error rates for getting it nearly right (e.g. predicted high, original medium) using the same calculation for error equal 40% (4+1+0+5=10/25). These errors are considered by this research to be errors, albeit not serious errors because the impact of providing a skin that is close but not right is less severe than providing a skin that is completely wrong. As such they are referred to as minor errors by this work. The error rate for getting it totally wrong (e.g. predicted high, original low) the error rate was 4% (1+0/25). These errors are referred to as serious errors. These results are summarised below in Figure 5-3:

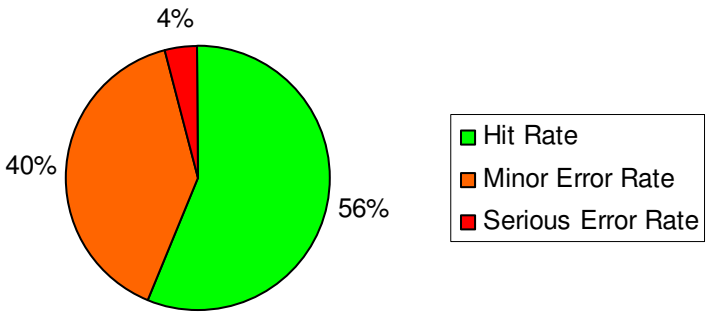


Figure 5-3 - Hit to Error Ratios including minor errors for Self Consciousness

5.5 Discussion

The results of the investigation provided evidence to support the experimental hypothesis, as demonstrated by an ability to predict personality trait from log file data.

As an exploratory study there were results that indicated links between personality and interactive behaviour. However, the low r^2 values from the regression analyses indicate that whilst personality is likely a factor influencing interactive behaviour, there are clearly other factors also involved. The results did not replicate the findings of Brebner and Cooper, Geen or Saati, Salem and Brinkman. However, there is evidence to suggest that personality is an individual difference that is expressed in interactive behaviours, and that these behaviours can be indicative of trait. Furthermore there is evidence to suggest that personality trait can be identified from interactive behaviour when categorised according to low, medium and high levels of a particular personality trait. However, the hit rate for the regression models in the crosstab analysis for an example trait, self consciousness, revealed 56% successes which are not particularly accurate to predict personality trait but may be of value in being able to identify personality trait from interactive behaviour when high degrees of accuracy are not required. An example application of such relatively low predictive success is personalisation of user interface skins, because it offers an ability for designers to provide an interaction that is more personally relevant than a single user interface. Such personalisation is enabled by any ability to understand the personality profile of the users and therefore even low accuracy methods for the determination of personality from log files provide a more personalised user experience than a single user interface approach.

5.6 Limitations

One of the limitations of this investigation was the low number of valid data sets available for analysis, despite the relatively high number of participants (143). As mentioned earlier this was due to conditions placed on participants as a means to encourage users to complete the IPIP-NEO questionnaire and to use the web radio. The consequences of this design decision was contrary to expectations and meant that users did not contribute both datasets required with a high frequency (approx 70%), resulting in low numbers of valid matching datasets. Future work should address this

issue. One possible solution is to make IPIP-NEO completion mandatory at installation time, rather than during web radio usage. If personality data was captured prior to usage then there seems like there is a higher likelihood of collecting both sets of data, because once the IPIP-NEO data has been collected all the user needs to do is to interact with the radio. This is in contrast to the method used in this research, which allowed radio usage first and then required IPIP-NEO completion during that time. It is suspected that this additional effort to complete IPIP-NEO whilst already being allowed to use the web radio was unreasonable for participants and did not encourage contribution of IPIP-NEO data.

It should be noted that the error rates referred to in this chapter are only intended as an example. In this particular example of self consciousness, the sample size is low at 25, out of a total of 115 participants (21.7%). This is explained by the lack of complete data sets for a large number of participants (90). As detailed in Chapter 3, participants were required to complete the IPIP-NEO questionnaire before being allowed to progress into the next phases of the experiment and receive skin packs. Some participants used the web radio but did not complete the IPIP-NEO questionnaire. Likewise, other users completed the IPIP-NEO but did not use the web radio. This conditionality placed on participants by the application had the effect of reducing their time using the application and as such did not contribute all the required data sets for this analysis, and hence the low number of valid cases in the crosstab analysis.

As a result of the imbalanced gender distribution the results and conclusions drawn from this experiment apply only males. Additionally, this experiment has focussed on an application-specific method for capturing interactive behaviour, with all behaviours specific to the WebRadio application. As such conclusions drawn should be viewed in the context of the WebRadio; however, future research in this field may reveal interactive behaviours that are indicative of personality trait in other applications that replicate results observed here. These may indicate norms or interactive behaviours that indicate personality trait across a broad range of applications.

5.7 Conclusions and Recommendations for Future Work

In conclusion, H_1 was weakly supported by the data collected in this experiment. However, under circumstances where high accuracy of prediction is required, for example replacing a psychometric inventory such as IPIP-NEO with a behavioural measure, according to this data, using a single dimension such as Personality may not provide the required accuracy. This is because the r^2 values from the regressions were relatively low and did not sufficiently account for enough variance in trait as a function of behaviour. For example, the IPIP-NEO correlates highly with the NEO-PIR with r values between 0.85 and 0.92, and the r values here ranged from 0.31 to 0.70. Had they been higher, stronger links would have been indicated between personality and interactive behaviour and most likely produced predictive models with greater r^2, and subsequent increased success in categorised personality trait testing resulting in greater accuracy. With typical r^2 values of 30% observed it seems likely that there is another factor or factors that are responsible for approximately 70% of the variance in personality trait that can be explained by interactive behaviour. Future work in this area could attempt to identify the other factors that contribute to any individual differences in interactive behaviour.

However, for less formal applications, such as designing personalised user interface skins, an ability to identify the personality profile of users however broadly is of benefit because it provides a more personalised experience and away from average – one of the stated aims of this research in Chapter 1. Under these circumstances it can be argued that there is evidence to support the experimental hypothesis from this data. The experimental hypothesis states that "Personality data can be extracted from log files". Even though the amount of variance in personality traits that was explained by interactive behaviour was low, there were still varying degrees of predictive success under both categorisation tests, with and without the minor error rates, for an example trait (self consciousness). Future work could use the concept of categorising personality into high, medium and low when testing predictive validity in attempting to include additional factors that might increase r^2 values and subsequently increase predictive ability of interactive measures.

Other behavioural metrics, such as keyboard button press speeds, could be explored to determine if there are interactive measures that can be defined from captured log file

data that are better predictors of personality trait than the ones defined by this experiment.

It was the intention of this research from the outset to investigate all 35 personality factors from the five OCEAN factors (Openness To New Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism) and to be able to predict Personality from these traits. However, this might have been too ambitious because perhaps certain personality factors are more clearly expressed in interactive behaviour than others at this level of interaction. For example, the artistic interests facet of openness to new experience may not be expressed at a micro level of interaction as clearly as a more macro level of interaction, like the number of visits to art galleries each year or number of photographs taken. As such there may be personality traits expressed in behaviour at a human computer interaction level that are more useful indicators of personality.

Having investigated the first experimental hypothesis relating to identification of personality, the next chapters address three different user interface design factors that can be modified to be more personally relevant to specific personality traits. The first of these is colour and is covered in the next chapter.

Chapter 6 – SKIN COLOUR PREFERENCE

6.1 Colour

This chapter addresses the second experimental hypothesis investigated by this research, that personality trait can predict user preference for colour in user interface skins. The chapter starts out by briefly reviewing the evidence supporting the experimental hypothesis. The next section describes the method used to investigate this particular hypothesis and includes material on how the skins were created, the two main datasets collected (self reported and behavioural) and the organisation and demographics of participants. The method section is concluded with an analysis methodology overview which describes the strategy for data analysis. The next main section is results and analysis, which provides summary output from all the analyses described in the previous section. Both the actual results and analysis are provided in the same section so that results and interpretation can be handled side by side for greater reader clarity. The chapter finishes off with the discussion section which provides observations and recommendations for future work.

A visual guide to the chapter is illustrated in Figure 6-1:

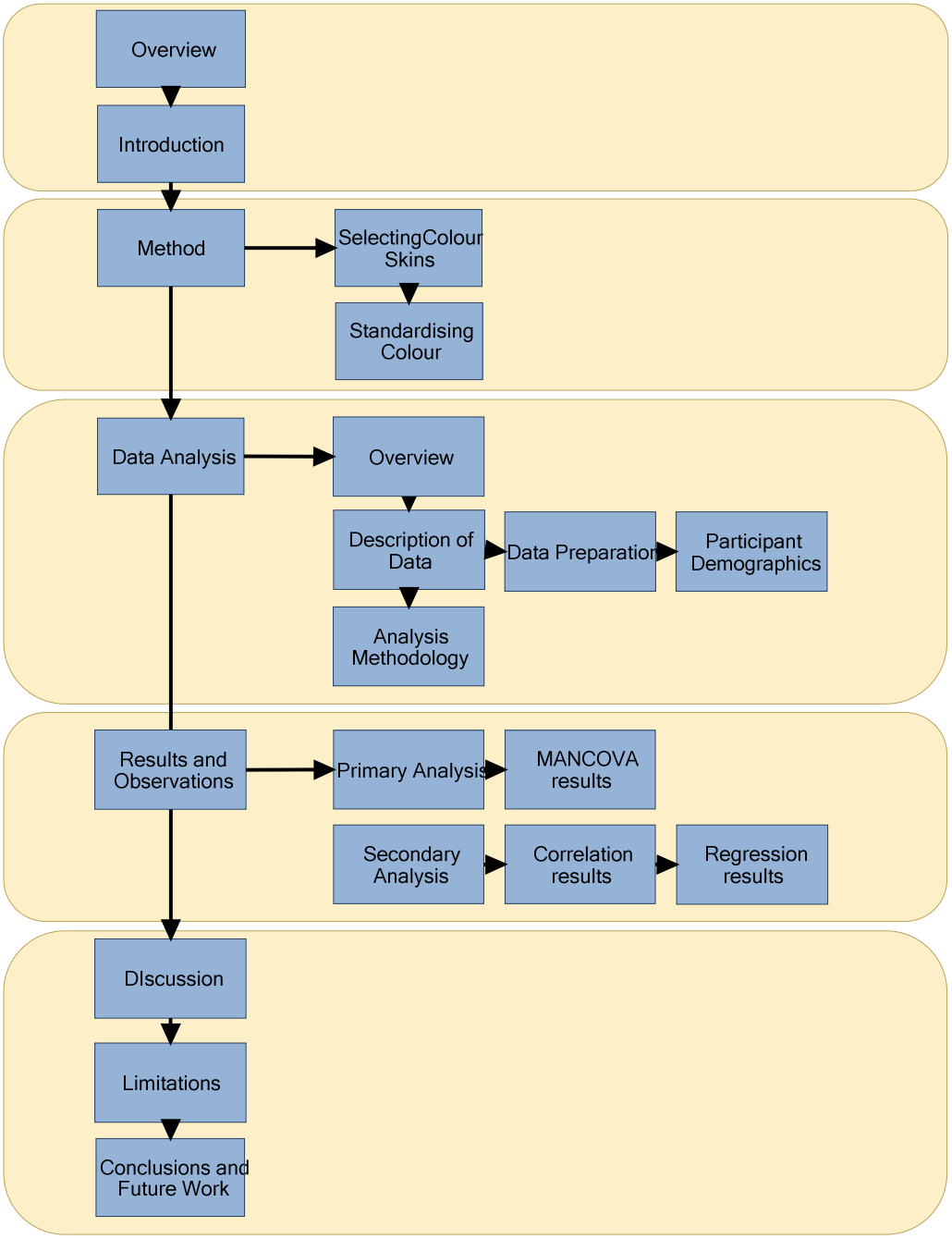


Figure 6-1 - Chapter Overview

6.2 Introduction

In the previous chapter the first experimental hypothesis was investigated to determine whether personality could be determined from user behaviour. This chapter investigates the second experimental hypothesis relating to interface skin colour:

H_2: User personality can predict preference for user interface skin colour

In trying to provide a more personally relevant interaction, this hypothesis attempts to link personality to user interface colour preference. The aesthetic affect of colour on human perception and the effects on interaction have been described in Chapter 2. To briefly summarise, links between personality and colour preference have been observed by Eysenck (1941), Barrett and Eaton (1947) who both observed correlations between personality trait and preference for paintings featuring either bright or subdued colours, with preference for brightness by extroverts and preference for subdued colours by introverts. This suggests that there is an individual response to colour as a function of personality. Furthermore there is evidence to suggest that this colour preference as a function of personality is operationalised by psychophysiological arousal (Eysenck, 1967, Johnson et al. 1999). Specific evidence supporting colour preference as a function of psychophysiological arousal includes research by Stone (2003), who observed task performance changing as a function of room colour, Valdez et al. (1994) who found that darker colours (blue, green, purple) were preferred to light colours (yellow, yellow/green) and that were also found to be the most physiologically arousing, Ali (1972) who observed greater arousal from red light than blue, Gerard (1957) who also observed greater arousal from red light than blue and Wolfson and Case (2000), who observed that game players performed better when exposed to blue screen colour than red.

As stated in Chapter 2, it is not within the scope of this research to investigate arousal itself as a factor. However it has been necessary to acknowledge it as an operationalised factor contributing to any relationships found between Personality and colour preference that could lead to a more personally relevant interaction investigated by this hypothesis. The hypothesised relationship between personality and colour preference is illustrated in Figure 6-2 below.

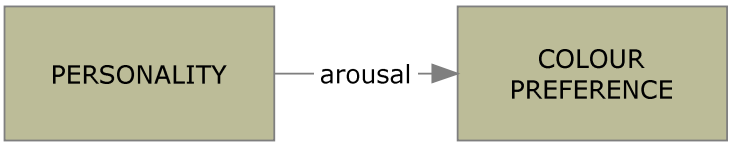


Figure 6-2 - Illustration of Hypothesis 2 factors, with personality predicting colour Preference and operationalised by psychophysiological arousal

This hypothesis investigates whether different personality traits influence user preference for colour in a user interface skin. The existing body of research suggests that brighter, more arousing colours will be preferred by extroverts, who according to Eysenck (1941) are inherently under aroused and seek an arousal equilibrium through a preference towards arousing stimuli (such as bright colours). Conversely, introverts who are inherently over aroused will seek to avoid arousing stimuli (such as darker colours). If this is the case then in this research it would be expected to observe similar preferences for light and dark colours by extroverts and introverts respectively.

6.3 Method

143 participants downloaded and installed the ProSkin WebRadio on their own personal computers in their own context, of which 64 provided viable data (55 male, 9 female). They completed a personality inventory (IPIP-NEO) and a skin preference questionnaire and the results stored on the central ProSkin database. Once the questionnaires were completed a package of seven different coloured user interface skins were uploaded to each client due to receive the colour skin package. The new coloured user interface skins were made available to the user and could be changed by the user by clicking on the 'setup' button. All user requested skin changes were recorded to log file and stored in the database. Both self reported skin preference data and behavioural data were collected so that comparisons could be made between the two.

6.3.1 Selecting Colour Skins

In order to investigate the relationships between user personality traits and user interface skin colour a set of coloured user interface skins were required. A set of colours for each of the main colours of the visible light spectrum were chosen: Red,

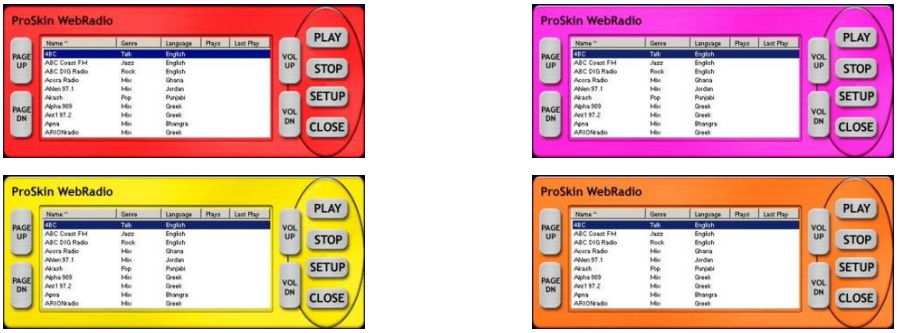
Yellow, Orange, Green, Blue, Purple and Pink. The colour Purple was chosen to replace and represent the colours Indigo and Violet which were shown to be more ambiguous to users in a pilot study than the other colours. The colour Pink was added because the pilot study also indicated a possible gender effect between females and Pink, and because it was not represented explicitly in the selected colours of the visible light spectrum.

An eighth colour skin, grey, is included in the analyses based on skin preference (self reported) data. It was not explicitly present within the colour skin pack but was made available to users along with the other coloured skins in the skin preference questionnaire, and therefore has been included. The data from the grey skin is not included in the behavioural data because it was not presented to users in the colour skin pack. The default grey skin was only available in initial phase, and not available in the colour skin package. Consequently no compariable behavioural data was available for the grey skin.

6.3.2 Standardising Colour

All colours were standardised to 50% saturation, 0% lightness and varied only for hue to determine colour so that all colours appeared at the same relative lightness and saturation to one another. No monitor calibration for colour temperature was conducted because the intended context of use would not necessarily have colour controlled monitors. However, by standardising for saturation and lightness the effect of any variance in colour representation is minimised because all colours appear relatively different to the user, removing the need to colour calibrate at the client side.

The final skins included in the colour skin pack are shown in Figure 6-3:



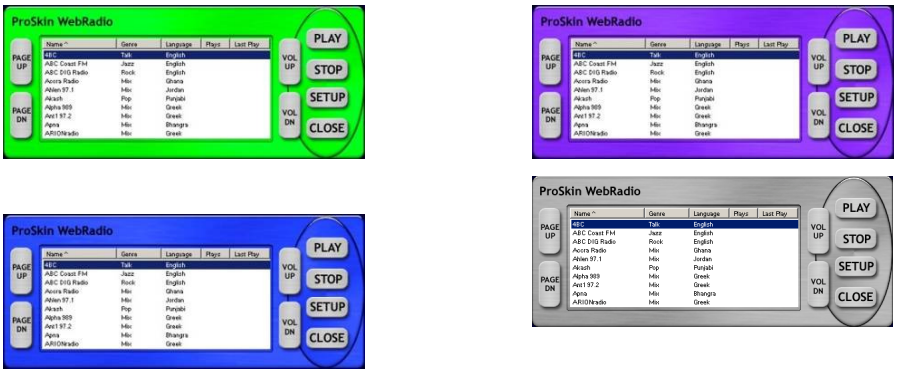


Figure 6-3 - Colour user interface skins used to test Hypothesis 2

The black bordering to each of the skins illustrated above does not appear to the user in the live user interface for the WebRadio. These images are the actual skins used in the WebRadio application and have a clipping colour applied that enables non-geometric shapes to be created, especially necessary for Hypothesis 3. When rendered by the WebRadio this clipping colour becomes transparent to the background. In the case of this hypothesis and this set of skins it enables rounded corners to these rectangles but for the shape hypothesis (H_3) enables complex shapes to be created.

6.4 Data Analysis

6.4.1 Overview

This section describes the data collected and the analyses performed to test the experimental hypothesis. Three datasets were collected for analysis - Personality, Skin Preference and behavioural interaction data. A description of these three data sets is supplied in the next section. Following on from the description and definition of the experimental data sets the next section describes the analysis methodology performed on the data. The actual results of the analysis and interpretation of them is then provided in a separate section.

6.4.2 Description of Data

6.4.2.1 *Self Reported Data*

Personality trait data of the participants was elicited using the IPIP-NEO personality inventory. This provided a set of 35 personality traits consisting of the five main OCEAN factors (openness, conscientiousness, extroversion, agreeableness, and neuroticism) each with seven sub factors. Trait scores were in the range 0 to 100, where 100 indicates a high level and 0 represents a low level of a particular trait. The trait score itself represents a percentile as compared with the general population. For example, in the case of extroversion there is a polar opposite,introversion. In this case a score of 0 represents low extroversion (i.e. high introversion) and 100 represents high extroversion. If a user were to score 72 for extroversion it would mean that 72% of the general population has a lower score (i.e. a percentile) and therefore mean that they would be classified as being in the top 28% for extroversion. Similarly for the Neuroticism factor a score of 100 represents high neuroticism and a score of 0 represents low neuroticism (also referred to as emotional stability). For the other factors, conscientiousness, openness and agreeableness there is no explicit adjective to represent low levels so for the purposes of this research are referred to as "non-" (e.g. non-conscientiousness) for the polar opposites.

All participant skin preference data was normalised to reduce the effect of any individual questionnaire answering strategy, for example always marking the polar extremes in responses or only using a small proportion of the scale. The method is described shortly in Section 5.4.3.1.

6.4.2.2 *Behavioural Data*

Any interaction with the ProSkin WebRadio application was recorded to a log file and transmitted to the central ProSkin database. The behavioural data was used as an indicator of user preference for user interface skin colour. This was done by extracting the amount of time that each user spent with each particular skin and then dividing that by the total amount of time spent viewing all the skins within the skin pack. This gives an indication of user preference by showing the proportion of time spent with each skin selected. The result of the formula gives a ratio of time spent on a particular skin as a percentage. The formula is represented below:

$$TimeSkinX = \frac{TimeSkinX}{\sum TimeSkin1...7}$$

Figure 6-4 - Formula used to indicate skin preference using behavioural data

The behavioural data collected for this hypothesis was very limited with a small number of people contributing data (n=9). This meant that there were not sufficient data recorded to perform the required multivariate analysis of covariance (MANCOVA) to test the experimental hypothesis. As a result of this low number the behavioural data set is used comparatively with the self reported data from the Skin Preference questionnaire where possible to determine any consistency.

6.4.3 Data Preparation

This section describes any preparatory work carried out on data prior to analysis. Firstly the self reported data from the skin preference questionnaire was normalised and secondly a reliability analysis was conducted to determine the internal consistency of skin preference data.

6.4.3.1 *Normalising Self Reported Data*

In order to normalise the self reported data from the skin preference questionnaire each response was compared with the participants own average response against their own deviation of rating skins. The formula for calculating this Z score is represented below in Figure 6-5, where x is the score to be normalised, is the mean rating given by an individual, and sd is the standard deviation of the individual's ratings:

$$z = \frac{x - \bar{x}}{sd}$$

Figure 6-5 - Formula for normalising skin preference scores

6.4.3.2 *Reliability of Self Reported Data*

User preference is by definition a highly subjective concept to investigate and therefore a number of additional questions were added to help increase the reliability of the preference data collected. This was done through the inclusion of additional "location" and "activity" items in the skin preference questionnaire. This enabled additional questions to be asked about preference for a particular skin, with the

intention of increasing the internal consistency. The questions consisted of four Likert scales with the following statements:

> "I would select this skin when working on my PC"
>
> "I would select this skin when working on my own"
>
> "I would select this skin when I am not alone"
>
> "I would select this skin when I am using the PC recreationally (e.g. family photos, playing games, using the WWW)"

A reliability analysis using Cronbach's Alpha to determine internal consistency was conducted that revealed high reliability for all colour skins, each with an alpha greater than 0.9. The alphas are summarised below in Table 5.6. General best practice dictates an alpha of 0.7 or higher (Sheskin, 1997) before a psychometric instrument can be considered internally consistent and so the alphas calculated for the four location items were assumed to be measuring the same construct reliably. This allowed skin preference questions to be analysed using an average of all four responses as an indicator of user preference for a particular colour skin.

Skin Colour	Alpha
Red	0.996
Yellow	0.993
Green	0.992
Blue	0.995
Pink	0.999
Orange	0.990
Purple	0.997
Grey	0.992
Mean	**0.994**

Table 21 - Reliability data for colour skin pack

6.4.4 Participant Demographics

6.4.4.1 *Gender*

Of the total 64 participants, 55 were male (85.9%) and 9 were female (14.1%), representing a large skew in the population. This is illustrated in Figure 6-6:

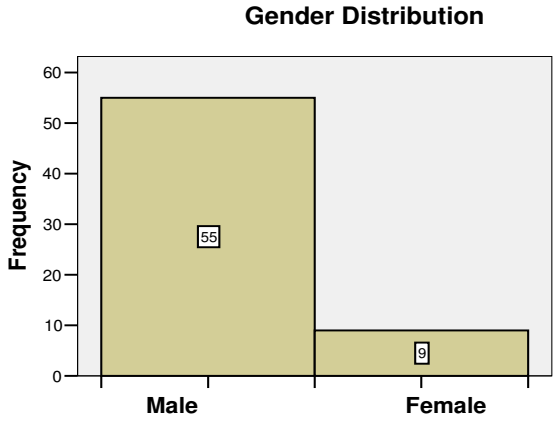


Figure 6-6 - Gender distribution for Hypothesis 2 (H_2)

6.4.4.2 *Age*

Although the experimental hypothesis does not directly relate to age as a factor in colour preference, a brief analysis was conducted to describe the age of the participant population. The external nature of the experiment (i.e. remote to the experimenter) meant that participants were not selected or seen by the experimenter, and so a descriptive view of the participant's age was performed.

The average age of participants was 33 years old, with a range of 46 years spanning the youngest at 19 years old and the eldest at 65. The standard deviation from the mean was 8.6 years. These statistics are summarised and illustrated in Table 22 and Figure 6-7:

N	Valid	64
	Missing	42
Mean		**33.47**
Std. Error of Mean		1.086
Median		**33.00**
Mode		42
Std. Deviation		8.689
Variance		75.491
Range		46
Minimum		19
Maximum		65

Table 22 - Descriptive Statistics for Age of participants contributing data to Hypothesis 2

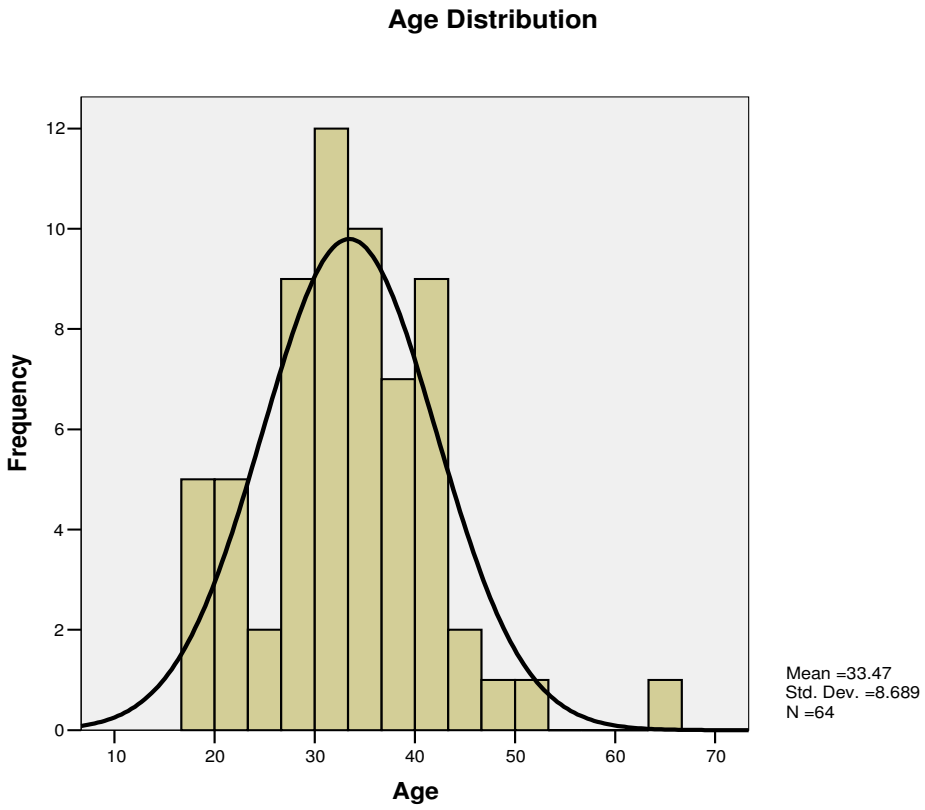


Figure 6-7 - Histogram with normalised curve showing age distribution for participants contributing data to Hypothesis 2

6.4.4.3 *Personality*

The personality trait data captured using the IPIP-NEO inventory was analysed using Kolmogorov-Smirnov tests for all five OCEAN personality traits and there was no indication that the data deviated from a normal distribution. The actual Z values from the Kolmogorov-Smirnov tests are summarised below in Table 23:

Trait	Z	sig
Extraversion	0.753	0.622
Agreeableness	0.649	0.793
Conscientiousness	0.480	0.975
Neuroticism	0.594	0.872
Openness	1.156	0.138

Table 23 - Kolmogorov-Smirnov Output for the OCEAN personality traits

One sample t-tests were performed for each of these IPIP-NEO traits examine whether this sample group was not significantly different when compared to a general population (50th percentile). All traits with the exception of Openness to New Experience were found to be not significantly different. These results are summarised below in Table 24:

	Test Value = 50					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Extraversion	-.650	63	.518	-2.016	-8.22	4.18
Agreeableness	.168	63	.867	.500	-5.45	6.45
Conscientiousness	-.771	63	.443	-2.469	-8.86	3.93
Neuroticism	.073	63	.942	.250	-6.55	7.05
Openness To New Experience	-3.322	63	.001	-10.891	-17.44	-4.34

Table 24 - One Sample t-test between OCEAN traits and the mean of 50% of the global population

Histograms are shown in Figure 6-8 to Figure 6-12 for each of the main factors to illustrate the personality profile of the participant group for H_2. The openness to new experience factor showed a small skew towards low openness to new experience participants, which is somewhat surprising given that participation in this research was voluntary. This suggests that the participants might have become involved through other motivators unrelated to new experiences such as wanting to help the experimenter out (possibly related to agreeableness).

Although openness to new experience was found to be significantly different there is still good range across the participants to get enough variation on the dimension to observe any effects. However, the lower openness scores for these participants as

compared to the general public means that the results of the analysis for this trait in particular should be considered with some reservation.

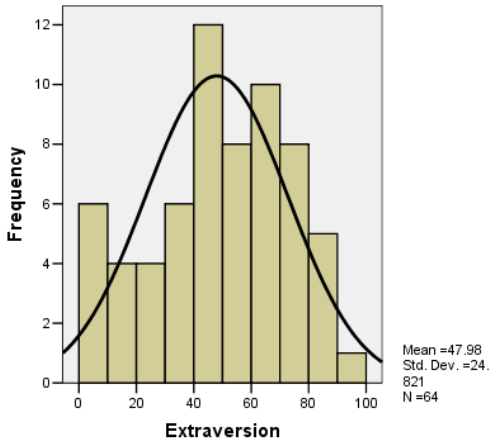


Figure 6-8 - Extraversion

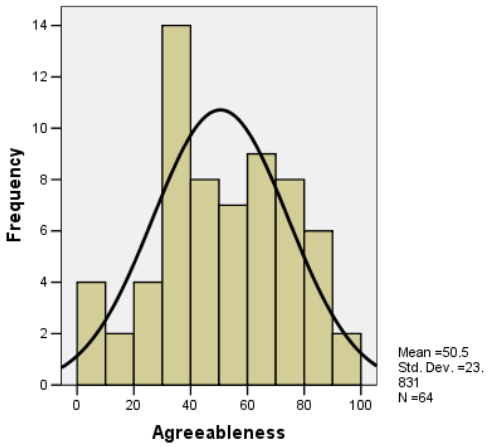


Figure 6-9 – Agreeableness

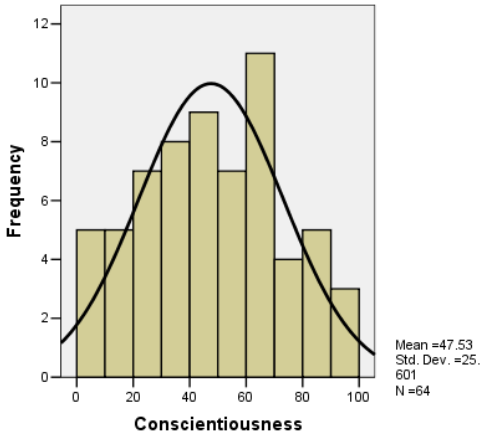


Figure 6-10 - Conscientiousness

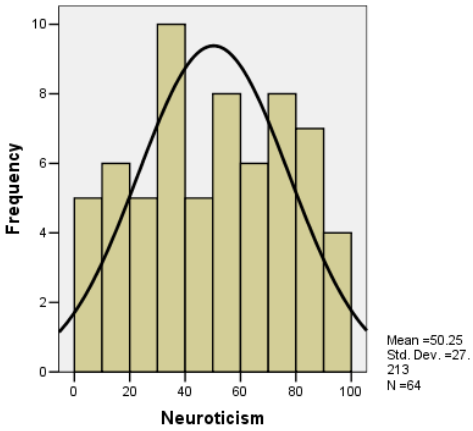


Figure 6-11 – Neuroticism

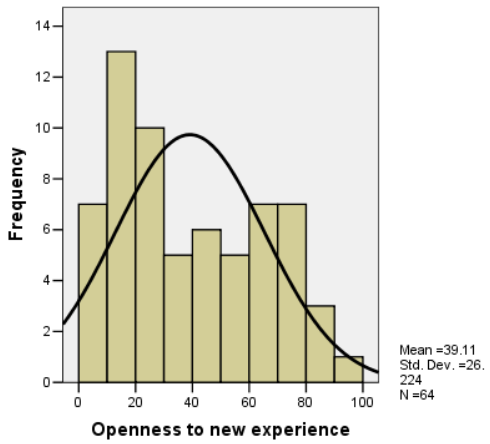


Figure 6-12 - Openness to New Experience

6.5 Analysis Methodology

Two levels of analysis were performed on the data collected. The primary, a priori analysis directly tests the experimental hypothesis and the secondary post hoc analysis provides an additional exploratory view of the remaining data relevant to the primary analysis. An overview of the methodology is provided below in Figure 6-13:

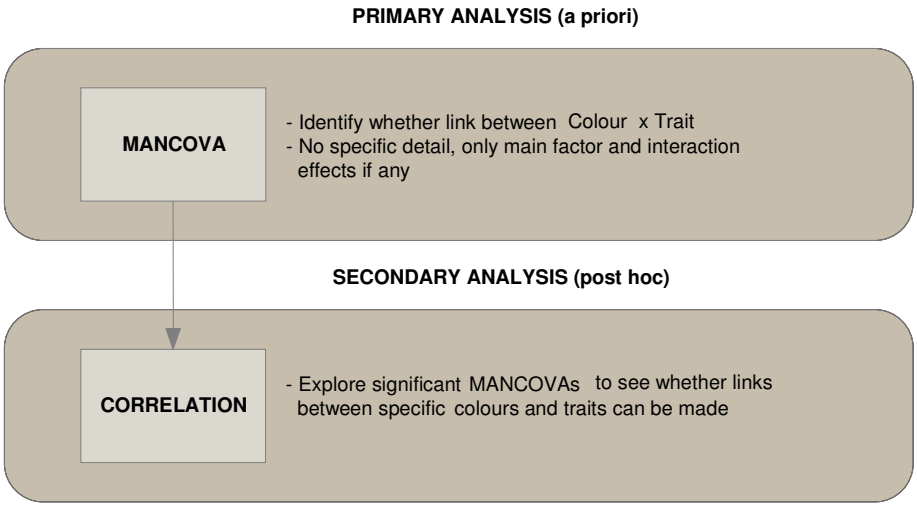


Figure 6-13 - Overview of Analysis Methodology

6.5.1 Primary Analysis

In order to test the experimental hypothesis (H_2) it was necessary to determine whether preference for user interface skin colour is a function of personality trait. The primary analysis looks at the effect of colour (in general) and personality trait and

any interaction that might exist between the two, in an attempt to establish a link between personality trait and user interface skin colour and thereby test H_2 and see whether personality trait can predict user preference for user interface skin colour.

Multivariate analysis of covariance (MANCOVA) was performed to determine what effect that colour and trait have in preference for user interface skin colour. This was done using a General Linear Model with repeated measures for two independent variables; the within-subject factor being Colour (eight levels: red, orange, yellow, green, blue, purple, pink, grey) and each of the 35 personality trait factors as the between subjects covariate (i.e. the MANCOVA was performed 35 times for each individual personality trait and sub factor). The dependent variable was the normalised skin preference scores for the self reported data.

6.5.2 Secondary Analysis

A secondary post hoc analysis was performed to try to see the effect of specific colours and different personality traits and has been performed on user skin preference data. The experiment generated a very large matrix of data consisting of the 35 IPIP-NEO personality traits and sub factors (for male, female and both genders) and both datasets from skin preference questionnaires and logged behavioural interaction data, so to run correlations between all of these factors would have been an undirected methodology that did not directly test the experimental hypothesis. However, as the data was already to hand and formatted for analysis this next level was performed on an exploratory basis only to see if anything could be learned for future work.

The purpose of the secondary data analysis was to explore the significant MANCOVA results that indicated a relationship between personality trait and colour (as factors in general). Two-tailed Pearson's correlations were performed between the significant personality traits from the MANCOVAs and both the skin preference and behavioural data, in an attempt to identify relationships between specific colours and traits (e.g. a link between extroversion and red). The correlations only highlighted relationships between traits and colours, so the correlations were then regressed and the function plotted on a graph to illustrate the observed correlations (e.g. extroverts prefer red).

This part of the investigation was performed as a post hoc analysis, using an adjusted alpha value to reflect the inherent increases to chance of committing a Type I error. This was calculated using the formula below (Sheeskin, 1997) in Figure 6-14, where represents the per comparison Type I error rate, represents the familywise Type I error rate and c represents the number of comparisons.

$$\alpha_{PC} = 1 - \sqrt[c]{1 - \alpha_{FW}}$$

Figure 6-14 - Formula for calculation of alpha for post hoc analysis of H_2

The number of comparisons performed in this experiment were eight, corresponding to the eight levels of colour (red, orange, yellow, green, blue, purple, pink and grey) and the familywise Type I error rate set at 5%. Calculating the new alpha using these values gives a post hoc alpha of 0.0064 shown below in Figure 6-15:

$$\alpha_{PC} = 1 - \sqrt[8]{1 - 0.05} = 0.0064$$

Figure 6-15 - Formula for calculation of alpha for post hoc analysis of Hypothesis 2 substituting for actual values

Having presented the analysis methodology, the next section presents the results along with a discussion.

6.6 Results and Observations

This section presents the results of the primary and secondary analyses.

6.6.1 Primary Analysis: Multivariate Analysis of Covariance (MANCOVA)

All participants, n=64												
	COLOUR				COLOUR AND TRAIT				TRAIT			
Trait	F	Hyp df	Error df	p	F	Hyp df	Error df	p	F	Hyp df	Error df	p
Excitement Seeking (E)	14.85	7	55	0.001**	3.933	7	55	0.002*	5.228	1	61	0.026*
Morality (A)	2.697	7	55	0.018*	3.423	7	55	0.004*	7.271	1	61	0.009*
Cooperation (A)	2.583	7	55	0.022*	5.139	7	55	0.001*	6.449	1	61	0.014*
Cautiousness (N)	3.428	7	55	0.004*	2.484	7	55	0.027*	2.551	1	61	0.115
Anger (N)	13.32	7	55	0.001**	2.394	7	55	0.033*	0.805	1	61	0.373
Liberalism (O)	2.469	7	55	0.028*	3.339	7	55	0.005*	0.048	1	61	0.827

Table 25 - Summary table for MANCOVA analysis on Skin Pack 2 (Colour), all participants

Male participants, n=55												
	COLOUR				COLOUR AND TRAIT				TRAIT			
Trait	F	Hyp df	Error df	p	F	Hyp df	Error df	p	F	Hyp df	Error df	p
Excitement Seeking (E)	18.18	7	46	0.001**	2.942	7	46	0.012*	11.48	1	52	0.001**
Morality (A)	4.528	7	46	0.001**	2.848	7	46	0.015*	10.85	1	52	0.002*
Co-operation (A)	2.526	7	46	0.028*	4.104	7	46	0.001**	14.24	1	52	0.003*
Liberalism (O)	3.874	7	46	0.002*	2.705	7	46	0.02*	0.023	1	52	0.881

*indicates significant at <0.05%, ** indicates significant at <0.01%,*

Table 26 - Summary table for MANCOVA analysis on Skin Pack 2 (Colour), male participants

Female participants, n=9												
Trait	COLOUR				COLOUR AND TRAIT				TRAIT			
	F	Hyp df	Error df	p	F	Hyp df	Error df	P	F	Hyp df	Error df	p
Trust (A)	337.5	7	1	0.042*	447.5	7	1	0.036*	2.942	1	7	0.130
Self Discipline (C)	211.5	7	1	0.053	409	7	1	0.038*	4.676	1	7	0.067

*indicates significant at <0.05%, ** indicates significant at <0.01%,*

Table 27 - Summary table for MANCOVA analysis on Skin Pack 2 (Colour), female participants

NOTE: These tables show only the significant results. The full output from all 35 traits and sub traits can be found in Appendix A.

The MANCOVA generates three F values, one for colour, trait and colour x trait (i.e. the interaction). Significant values for colour indicate that colour is a factor determining skin preference. The results confirm what one would expect, that colour on its own is a factor in user interface skin preference. The trait values represent preference for skin packages in general (i.e. colour, shape, meaning). To illustrate this, Figure 6-16 below shows two examples of preference for skin package as a function of trait (in this example morality and extroversion).

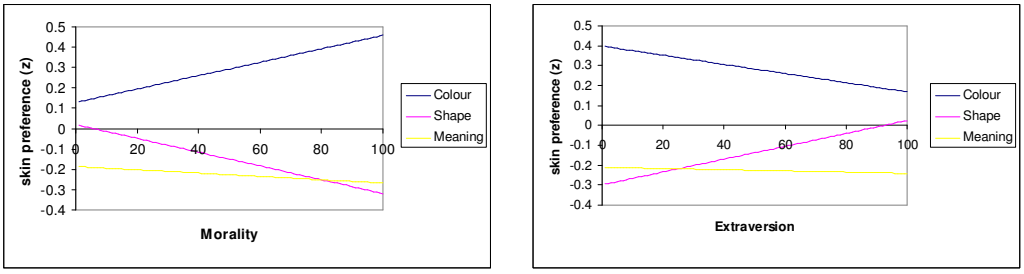


Figure 6-16 - Examples of two traits to illustrate preference for skin pack type as a function of personality trait

The values for user preference for skin pack are interrelated to each other because during the normalization process all scores are made relative to the mean for all skins. For this reason these graphs show relative preference but does not show causation by a specific trait. However they are presented here to illustrate what the trait output from the MANCOVA means, that there are differing preferences for the skin pack types as a function of personality trait.

Significant values for trait indicate that personality is a factor in determining preference for skin pack content (i.e. preference for colour skins over shape or meaning skins). The interaction of these two factors, represented by colour x trait, if significant indicates that both colour and trait combined is a significant factor in determining skin preference. These experimental variables are illustrated below in Figure 6-17. The analyses were performed for males and females separately and the results recorded (see Table 25, Table 26 and Table 27).

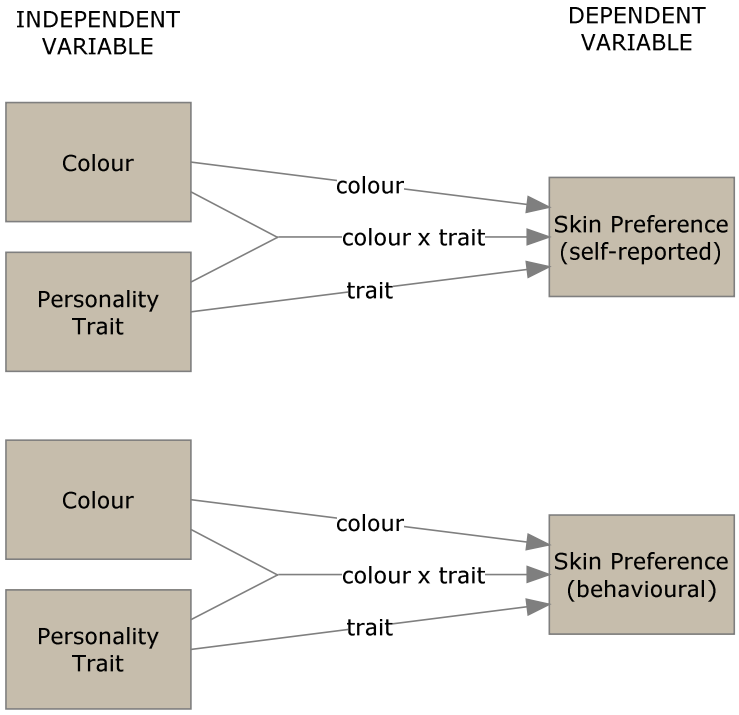


Figure 6-17 - Experimental variables for Hypothesis 2 showing the three experimental effects consisting of two main effects (Colour, Trait) and one interaction effect (Colour x Trait) for both self reported and behavioural data

The results of the MANCOVAs for colour and trait show significant interaction effects between colour and personality trait on skin preference for six of the thirty five IPIP-NEO personality traits. This indicates that for these particular traits, excitement seeking, morality, cooperation, cautiousness, anger and liberalism, that personality trait does affect preference for user interface skin colour. Four of the five main OCEAN factors were represented with only conscientiousness not showing any significant results.

These findings confirm the hypothesis that personality trait has an effect on users preference for colour in user interface skins.

6.6.2 Secondary Analysis

This section provides a discussion of the results of the secondary or post hoc analysis. The MANCOVA analysis previously conducted in the a priori analysis confirmed that the interaction effect of colour and trait can influence user interface skin colour preference. The purpose of performing these post hoc correlations is to try to identify whether relationships exist between specific colours, and if so to determine which direction the correlations are in. Data was analysed for all participants and then by gender to determine if there were any gender related differences. All alpha levels for this analysis were performed using the post hoc alpha level of 0.0064 as defined in section 5.5.2.

	All Participants	Skin Colour	Self Reported	p (two tailed)	Behavioural	p (two tailed)
Trait	Factor		n=63		n=10	
E	Excitement Seeking	Grey	-0.425	0.001		
		Yellow	-0.364	0.003	0.171	0.829
A	Morality	Grey	0.388	0.002		
		Purple	0.42	0.001	0.495	0.670
A	Cooperation	Grey	0.468	0.001		
C	Cautiousness	Grey	0.445	0.001		
N	Anger	Grey	-0.379	0.002		

Table 28 - Significant Pearson Correlations between Trait and Skin for all participants (n=63)

	Male Participants	Skin Colour	Self Reported	p (two tailed)	Behavioural	
Trait	Factor		n=54		n=8	
E	Excitement Seeking	Grey	-0.398	0.003		
		Yellow	-0.436	0.001	0.171	0.829
		Purple	0.383	0.004	-0.863	0.337
A	Morality	Grey	0.391	0.003		
		Purple	0.456	0.001	0.495	0.670
A	Cooperation	Grey	0.453	0.001		
C	Cautiousness	Grey	0.417	0.002		
N	Anger	Grey	-0.397	0.003		

Table 29 - Significant Pearson Correlations between Trait and Skin for male participants (n=54)

Trait	**Female Participants** Factor	Skin Colour	Self Reported n=9	p (two tailed)	Behavioural n=2
C	Self Discipline	Green	-0.857	0.003	
O	Trust	Red	-0.848	0.004	

Table 30 - Significant Pearson Correlations between Trait and Skin for female participants (n=9)

There are correlations omitted from the behavioural data above in Table 29 and Table 30. These are the correlations for the grey skin which were not presented to the participants at the same time as the colour skins. This was because the grey skin was part of the default skin pack provided at the time of installation and was not an option for users to select when in the colour skin pack experimental condition. This meant that the behavioural data, calculated as the ratio of time a particular skin was used as a function of the entire skin pack, could not be compared with the self reported data in a meaningful manner. As a result of this it was not possible to include these results in the table. Furthermore, the low number of female data available for analysis (n=2) meant insufficient data for this statistical test and the results omitted.

These issues are discussed in greater detail in the section decribing limitations

6.7 Summary of Correlations

The following graphs summarise and illustrate the correlations (Figure 6-18). A full summary table with all correlations within 5% a priori alpha is provided in the Appendix B for reference. For example, the graph for excitement seeking shows two significant correlations for the grey and yellow skins. This is a negative correlation and so it shows that as the user's excitement seeking score increases, preference for the grey and yellow skins decrease.

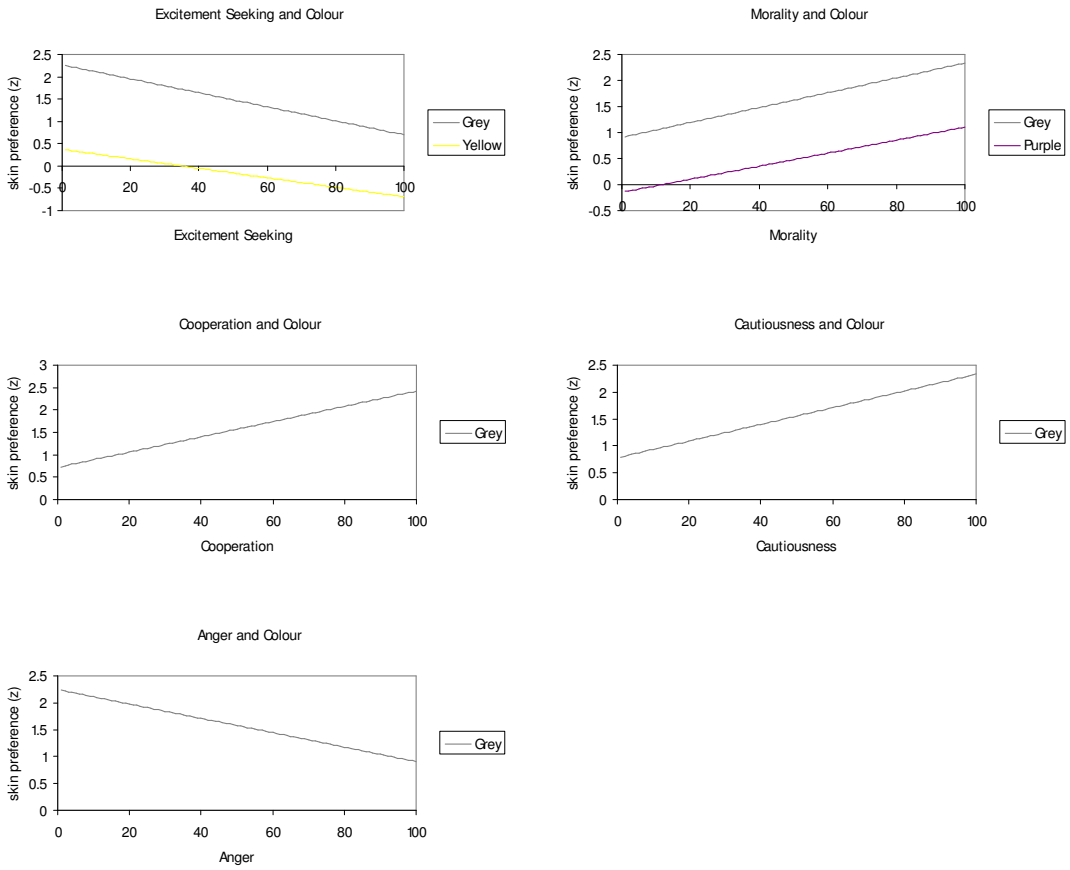


Figure 6-18 - Illustration of significant correlations (all participants, n=63 under post hoc alpha = 0.0064)

All significant correlations under the experimental post hoc alpha of 0.0064 were almost the same for the male population as for both genders. This is most likely due to the large difference in gender distribution, with a high number of male participants (54) and a low number of female participants (9). As a result of this difference the subsequent regression analysis focuses solely on the data from male participants (see section below on gender).

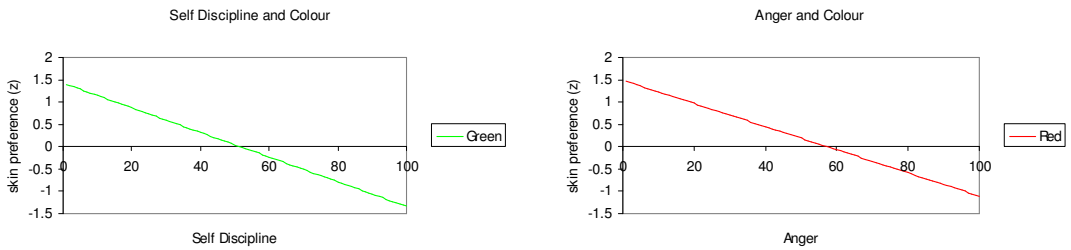


Figure 6-19 - Illustration of significant correlations (female participants, n=9 under post hoc alpha = 0.0064)

6.7.1 Regression Analysis

Using a stepwise method regression models of user interface skin colour preference were created. The parameters entered in the stepwise method were each of the significant traits from the correlation analysis. This meant that for grey there were 5 traits, for purple there were 2 traits and for yellow 1 trait entered into the respective models. The results are summarised below in Table 31:

Colour	r	r^2	adj r^2	F	Df	sig	Predictor	B	t	sig
Grey	0.52	0.27	0.24	9.23	2	0.001	Cooperation	0.01	2.52	0.015
							Cautiousness	0.01	2.05	0.046
Yellow	0.44	0.19	0.17	12.21	1	0.001	Excitement Seeking	-0.01	-3.49	0.001
Purple	0.46	0.21	0.19	13.66	1	0.001	Morality	0.01	3.70	0.001

Table 31 - Regression models of skin colour preference by trait for each of the skins with significant trait correlations (males, n=54).

The result of the analysis indicates that grey is the colour that most successfully explains the variance for skin colour preference, accounting for approximately 27% of the variance. Two factors increased the accuracy of this model, cooperation and cautiousness. The next most successful skin colour was purple, accounting for approximately 21% of the variance with yellow accounting for approximately 19%.

These findings suggest that personality trait data can be used to partially account for skin colour preference, providing evidence to support the experimental hypothesis.

6.8 Discussion

The findings of this study support the experimental hypothesis, that personality trait can partially predict user preference for colour in user interface skins. The results from the a priori MANCOVA indicated that personality trait is a factor for colour preference in user interface skin. The post hoc investigation identified a number of correlations between specific colours and personality traits, which when regressed were able to help form predictive models for user interface skin colour. The regression models at best were able to account for approximately 27% of the variance.

These findings did not replicate the work of Brinkman, Saati and Salem (2005), who found predominantly found correlations with the blue user interface skin. No correlations were observed for either sex for the blue skin. Brinkman et al. did find a positive correlation with the yellow user interface skin, however it was for the achievement striving facet of conscientiousness and the only correlation observed in this study for conscientiousness was the grey skin.

Furthermore, no preference for brighter colours (previously suggested as more arousing) was observed by extroverts, as in the Eysenck (1964) studies. The only significant brightly coloured skin was the yellow skin and the direction observed was opposite to the expectation with preference for the yellow skin decreasing as a function of extroversion.

6.9 Limitations

There were a number of limitations that should be noted. There was an obvious gender imbalance with 85.9% males and 14.1% females. This imbalance heavily influenced data for all participants and meant that there was too little data for the analysis of female data. As a result of this the decision was taken to focus on the data from the male participants, where there was sufficient data to perform the analysis. This means that the conclusions drawn from the analyses here are only relevant to males and there is no comparison or analysis for females.

The limited amount of behavioural data collected meant that there was not enough data to perform a statistical analysis. This was due to the conditions applied to each

user of the WebRadio application, which determined that the user must complete the IPIP-NEO personality inventory and skin preference questionnaires first before proceeding to download new skin packs. This meant that there was a tendency for users to stay in the default condition (two skins, grey and a helpskin) and not progress into the colour skin pack phase. As a result of this very few people (n=9) contributed behavioural data for colour skins which was too limited for use.

At the outset of this chapter, research into the effect of psychophysiological arousal on colour preference was described. The research suggested that lighter colours were possibly responsible for greater arousal than subdued colours. The results from this experiment did not provide clear evidence to support this underlying mechanism. For example, in Figure 6-18 two correlations can be seen for Excitement Seeking for both the grey and yellow skins. The correlation for grey shows a decrease in preference as a function of excitement seeking. One interpretation of this finding is that grey may be considered a boring or a dull colour to people with excitement seeking tendencies, and therefore was less likely to be preferred. According to the previous research, one would have expected a lighter more arousing colour like yellow to have been preferred by extroverts (the main trait for the excitement seeking facet), however as can be seen in Figure 6-18, the results from this study showed yellow to be preferred less than grey for extroverts. This suggests that either the levels of arousal were not sufficient to observe the expected phenomenon, or that perhaps there are other factors involved in skin preference in addition to arousal. It is possible that the actual colours selected for the experiment were not sufficiently arousing to cause a change in behaviour; however there were no observations made during the experiment to support this idea.

6.10 Recommendations for Future Research

There are a number of recommendations arising from this experiment. The first and most obvious one is the addressing of the gender imbalance to both provide greater insight into female preference for colour, and for comparison with the male data.

The actual colours chosen by the experimenter were picked arbitrarily and then standardised for saturation. Perhaps future work could investigate preference for specific colours to determine which particular properties of a colour are preferred (for

example, RGB or CMYK). It might be the case that the experimental colours chosen were either not arousing enough, or even just not appealing enough. Investigation into specific shades of a particular colour might help highlight better examples of specific colours to use in an experiment like this, which looks at examples from each colour of the visible spectrum. Another factor to be considered for future research is the size of the user interface compared to the size of the available screen space, i.e. the ratio of user interface to background. It is possible that more colour space visible to the user may increase the effect of colour on preference.

The limited amount of behavioural data, and the conditionality applied to participants that caused it, could be revised in future research to remove the behavioural data investigation altogether. This research used a single experimental platform and data collection period for all four experimental hypotheses. In so doing it might have unnecessarily restrictive conditions were applied to participants and this limited the amount of data to the point at which data was unusable. The solution might be to use the experimental platform in a separate phase to investigate behavioural data versus self reported data.

6.11 Conclusion

This chapter has investigated the second experimental hypothesis and found that according to the experimental data, personality is a factor influencing user preference for colour of user interfaces. The analyses revealed sufficient evidence to accept the experimental hypothesis. Additional analyses were performed to try to learn more about specific colours, but gender imbalance meant that only male participant data was used. The analyses revealed relationships between three coloured skins and a number of personality traits, ranging between approx 19 and 27% also providing support for the experimental hypothesis – that personality trait can partially predict preference for skin colour. By designing for different personality traits a more personally relevant interaction can be provided, the overall aim of this research.

The next chapter investigates the third experimental hypothesis which investigates the relationship between personality and user interface shape.

Chapter 7 – SKIN SHAPE

7.1 Shape: Introduction

This chapter continues the hypothesis testing with the investigation of the third experimental hypothesis relating to interface skin shape. Both this and the previous hypothesis investigate individual responses to interface design features (colour, shape) that are operationalised by psychophysiological arousal.

The chapter starts out with a description of the method used to select the shape skins to be used in this investigation and then the methodology used to collect the data. This hypothesis was tested using the same experimental methodology as for H_2 and therefore only the key points are repeated from the previous chapter for the sake of structure and ease of reference. The results of the analyses are then presented and conclusions drawn. No evidence to support this hypothesis was found during the analysis of this dataset. A visual guide to the chapter can be found in Figure 7-1.

H_3: User personality can predict preference for user interface skin shape

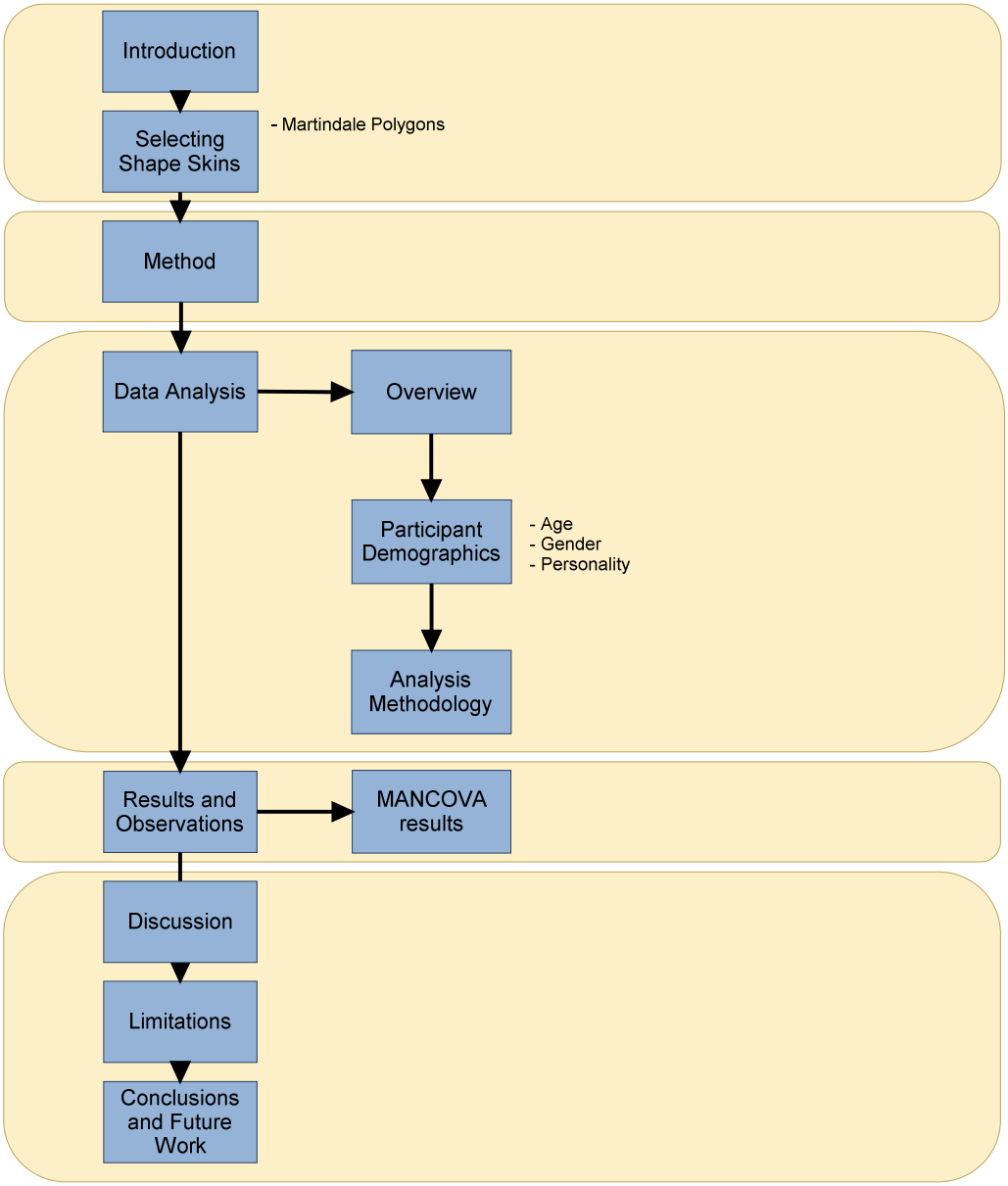


Figure 7-1 - Visual guide to Chapter 6

7.2 Selecting Shape Skins

In order to investigate any potential relationships between user personality traits and user interface shape a set of differently shaped user interface skins were required. The literature was searched for examples of shapes used in personality research and an area of research regarding random polygon shape and arousal in extroverts was found. Before providing the details of the actual skins selected for the experiment, a brief summary of the key literature points that determined the skins now follows.

Early research in this field was conducted by Munsinger and Kessen (1964) who investigated the effect of visual complexity on arousal. They used a set of randomly sided polygons with differing numbers of turns ranging from 5 to 20. They found that the relationship between complexity and arousal was represented by an inverted-u shape, or Wundt curve, when expressed in user preference. Berlyne (1972) supporting the Munsinger and Kessen findings of a Wundt curve, suggested that preference for any stimulus is determined by arousal potential. However, Martindale, Moore and Borkum (1990) conducted a series of experiments using randomly shaped polygons to further investigate the findings of Berlyne and found that additional factors contributed to preference, such as size and meaningfulness (the meaningfulness component is investigated in H_4 in the following chapter). They did however find that in some cases that complexity (defined as number of turns in a randomly shaped polygon) and preference were represented by a u-shape.

Both the Munsinger and Kessen (1964) and Martindale et al. (1990) studies used randomly shaped polygons controlled for number of turns (5, 6, 8, 10, 13, 16) so it was decided to replicate their methodology by using similarly differentiated polygons. A set of shapes were generated randomly using 5, 7, 10, 13 and 20 turns to replicate the original 1964 and 1990 polygons and an additional two were added, 30 and 40 turns to provide a more extreme stimulus to test the hypothesis with. The actual polygon shapes were not replicated, just the number of turns. The actual skins used in this skin pack are illustrated in Figure 7-2:

Although there are additional factors that describe shape, such as symmetry, curvature of sides and variance in size depth, this research investigates one factor of shape as defined by Munsinger and Kessen, and by Martindale et al. This is because there is existing literature from which to investigate. Suggestions for future work in this field includes investigation into different shape factors and can be found in Section 6.8.

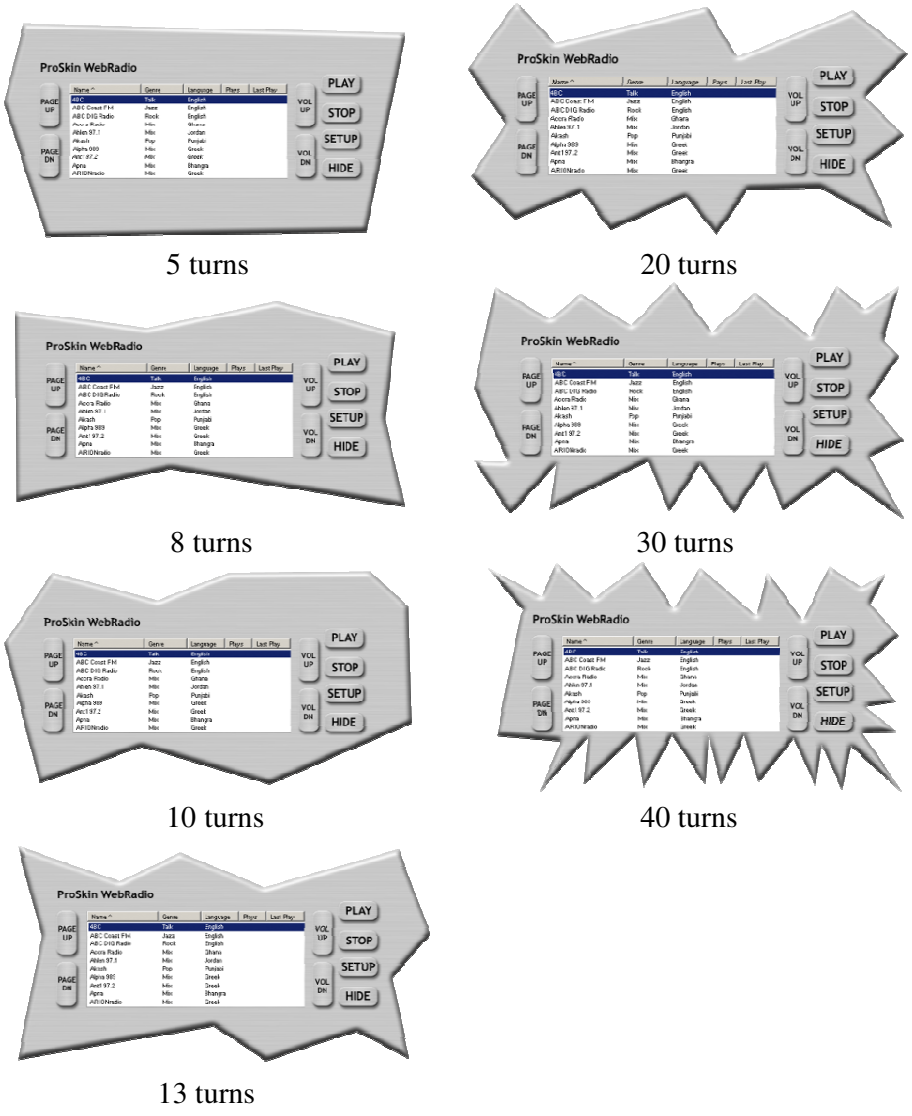


Figure 7-2 - Skins used to test Hypothesis 3 showing random polygons using number of turns from previous studies (5, 8, 10, 13, 20) and two created for this experiment (30 and 40)

7.3 Method

A total of 143 participants downloaded and installed the ProSkin WebRadio on their own personal computers in their own context, of which 64 provided viable data (55 male, 9 female). They completed a personality inventory (IPIP-NEO) and a skin preference questionnaire and the results stored on the central ProSkin database. The skin preference questionnaire contained images of all the skins from the three skin

packs (colour, shape and meaning) and responses to preference questions placed on a seven point Likert scale. Once the questionnaires were completed a package of seven different shaped user interface skins were uploaded to each client. The new shaped user interface skins were made available to the user and could be changed by the user by clicking on the 'setup' button. All user requested skin changes were recorded to log file and stored in the database. Both self-reported skin preference data and behavioural data were collected so that comparisons could be made between the two.

7.4 Data Analysis

7.4.1 Overview

This section describes the data collected and the analyses performed to test the experimental hypothesis (H_3). As for H_2, two datasets were collected for analysis – self reported Personality data as measured by IPIP-NEO, and the logged behavioural data from the same participants as H_2. The description of these datasets can be found in Chapter 5. As for the colour skin preference questionnaire responses, the shape responses were also normalised to minimise the effect of personal answering strategy.

7.4.2 Reliability Analysis

As for the previous hypothesis, reliability analysis was conducted on responses to four questions each addressing the same underlying construct relating to context of use. This allowed a single question to be asked of participants that represented their preference for skin shape irrespective of the context of use. The results from the Cronbach's Alpha analysis can be found in Table 32:

Skin Shape	Alpha
# turns	
5	0.999
7	0.995
10	0.988
13	0.944
20	0.966
30	0.989
40	0.992
Mean	0.982

Table 32 - Summary of reliability analysis data

7.4.3 Participants' Demographics

The participant population for H_2 was the same for H_3, 55 males and 9 females of average age 33 (SD = 8.68). Details of age, gender and personality trait data can be found in section 5.4.4.

7.5 Analysis Methodology

The same analysis methodology for colour (H_2) was adopted to investigate shape (H_3), involving a primary analysis using MANCOVA to determine effect of shape and trait in preference for user interface skin. Full details can be found in Chapter 5. The dependent variable was the normalised skin preference scores from the shape skins. The two independent variables were the user's personality trait and the shape of the skin.

7.6 Results and observations

This section presents the results of the data analysis for this hypothesis. The results of the MANCOVA can be seen in Table 33 and observations made underneath.

Trait	SHAPE				SHAPE and TRAIT				TRAIT			
	F	Hyp df	Error df	P	F	Hyp df	Error df	p	F	Hyp df	Error df	P
Extra-version	1.964	6	56	0.086	0.815	5	56	0.563	1.816	1	61	0.183
Friendliness	2.372	6	56	**0.041***	1.610	6	56	0.161	0.975	1	61	0.327
Greg-ariousness	1.659	6	56	0.148	0.799	6	56	0.575	1.578	1	61	0.214
Assert-iveness	1.070	6	56	0.391	0.296	6	56	0.936	0.507	1	61	0.479
Activity Level	2.706	6	56	**0.022***	0.504	6	56	0.803	1.729	1	61	0.193
Excitement Seeking	2.366	6	56	**0.042***	0.629	6	56	0.706	0.003	1	61	0.960
Cheerfulness	2.590	6	56	**0.028***	1.018	6	56	0.423	0.030	1	61	0.864
Agree-ableness	0.610	6	56	0.722	0.653	6	56	0.687	3.406	1	61	0.070
Trust	1.001	6	56	0.434	0.734	6	56	0.625	0.421	1	61	0.519
Morality	0.724	6	56	0.632	0.196	6	56	0.977	2.576	1	61	0.114
Altruism	1.697	6	56	0.139	0.394	6	56	0.880	4.886	1	61	**0.031***
Co-operation	0.773	6	56	0.594	1.311	6	56	0.267	2.916	1	61	0.093
Modesty	1.364	6	56	0.245	1.019	6	56	0.422	0.166	1	61	0.686
Sympathy	1.447	6	56	0.213	0.430	6	56	0.856	0.520	1	61	0.473
Conscien-tiousness	1.387	6	56	0.236	0.685	6	56	0.662	0.005	1	61	0.946
Self Efficacy	3.282	6	56	**0.008****	1.256	6	56	0.292	0.263	1	61	0.610
Orderliness	1.762	6	56	0.124	0.764	6	56	0.602	0.206	1	61	0.652
Dutifulness	0.897	6	56	0.504	0.795	6	56	0.578	1.767	1	61	0.189
Achievement Striving	2.730	6	56	**0.021***	0.736	6	56	0.623	0.023	1	61	0.880
Self Discipline	1.999	6	56	0.081	0.707	6	56	0.645	2.912	1	61	0.093
Cautiousness	0.549	6	56	0.768	0.914	6	56	0.492	0.058	1	61	0.810
Neuroticism	2.237	6	56	0.053	1.126	6	56	0.360	0.523	1	61	0.472
Anxiety	2.156	6	56	**0.008****	1.728	6	56	0.131	0.140	1	61	0.710
Anger	3.118	6	56	0.100	0.724	6	56	0.632	0.098	1	61	0.756
Dep-ression	1.937	6	56	0.091	1.182	6	56	0.329	0.131	1	61	0.719
Self Cons-ciousness	1.123	6	56	0.361	0.517	6	56	0.793	3.046	1	61	0.086
Immod-eration	1.557	6	56	0.177	0.403	6	56	0.874	1.018	1	61	0.317
Vulnerability	1.063	6	56	0.396	0.924	6	56	0.485	0.684	1	61	0.411
Openness	2.543	6	56	**0.03***	0.747	6	56	0.614	3.994	1	61	0.050
Imagination	3.750	6	56	**0.003****	1.548	6	56	0.180	4.048	1	61	**0.049***
Artistic Interests	1.615	6	56	0.160	0.650	6	56	0.690	2.349	1	61	0.131
Emotionality	2.500	6	56	**0.033***	0.630	6	56	0.706	0.857	1	61	0.358
Advent-urousness	2.887	6	56	**0.016***	1.397	6	56	0.232	0.015	1	61	0.902
Intellect	1.617	6	56	0.159	0.529	6	56	0.784	0.739	1	61	0.393
Liberalism	1.025	6	56	0.419	0.801	6	56	0.573	0.728	1	61	0.397

**indicates sig <0.05%, ** indicates sig at <0.01%*

Table 33 - MANCOVA analysis results from Shape skins for all participants (n=57)

Multivariate analysis of covariance (MANCOVA) did not provide any results to support the experimental hypothesis because no significant interaction effect was found. The two main effects (shape, trait) will be examined first before looking at the interaction effect of shape x trait.

The main effect of shape shows 10 cases with significant effects, meaning that the shape of the interface was a factor that determined skin preference for those traits. However, it does not indicate any direction for the preference. In order to achieve this, an MANOVA with repeated measures was performed with only shape entered as the within subjects independent variable. The results confirmed a significant effect for preference as a function of shape ($F(6,57)=4.78$, $p=0.01$). These are illustrated below in Figure 7-3 and Figure 7-4 for males and females respectively.

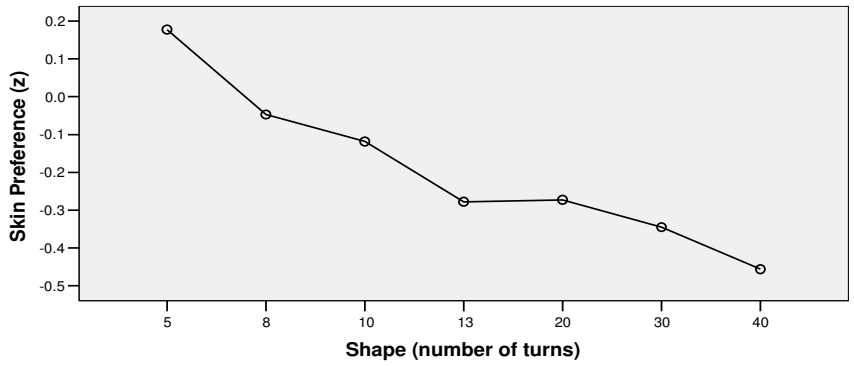


Figure 7-3 - Plot of mean preference for shape as a function of the number of turns (male participants, n=54)

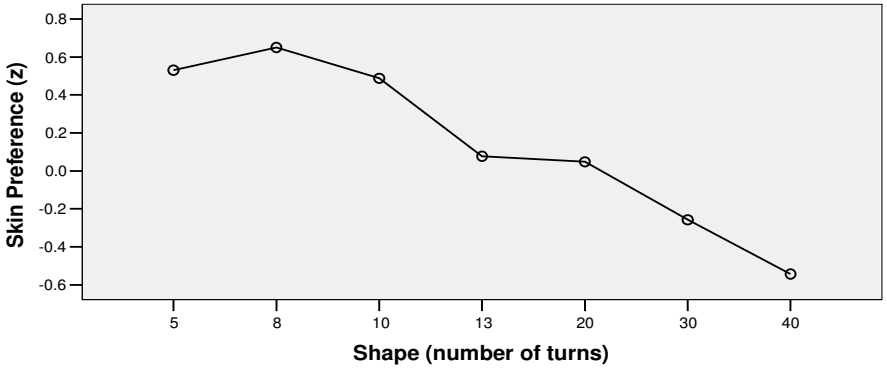


Figure 7-4 - Plot of mean preference for shape as a function of the number of turns (female participants, n=9)

Unlike colour, the number of turns in a polygon can be considered ordinal data, or even interval data, and as such affords comparative analysis. For example, a polygon could be described as becoming more complex as the number of turns increases, with that complexity a function of the number of turns. The same cannot easily be said for colour which for general use is considered nominal data. Figure 7-3 and Figure 7-4 above show a tendency for a lack of preference increasing as a function of number of turns.

7.7 Discussion

The findings from this experiment run contrary to expectation because Berlyne's work suggests that optimal arousal would influence skin preference, and that this would be represented by a Wundt curve in an inverted u-shaped manner. These results do not support Berlyne's work and there are a number of possible reasons for this. Firstly the arousal levels caused by the variance in the experimental shapes were not sufficient to observe a Wundt curve. Secondly arousal may not be the dominating factor affecting skin preference with other factors responsible. Thirdly, the experimental polygons used in this experiment were applied as user interface shapes, whereas in Berlyne's experiments these polygons were presented as shapes. Preference for shape in general may well be different to preference for interface shape – when "embodied" as a user interface the shape of an object can take on different affordances. For example, an irregular shaped polygon does not use space optimally when in the context of a desktop application, whereas a regular shaped polygon typically does. These limitations are discussed in the limitations section next. In this experiment increasing polygon complexity in a user interface was not preferred by participants.

The main effect of trait indicates different people's preference for different types of skin packs. There were only two significant results found for altruism (a facet of agreeableness) and imagination (a facet of openness). Correlations were performed on these facets to determine the direction of any preference for the skin shape pack in general. Results showed negative correlations for altruism (r=-0.272, sig (2-tailed)=0.031) and imagination (r=-0.249, sig (2-tailed)=0.049) indicating a lack of preference of the shapes skin pack as a function of these two traits.

One interpretation of the significant results for altruism is that participants with highly altruistic personality traits were acting more selflessly and acquiescing to what they thought were the intentions of the research, and this should be noted as a potentially confounding variable. As for the significant result for imagination, one interpretation is that more imaginative individuals did not like the irregularly shaped polygons comprising the shape skin pack because they were asymmetrical and not aesthetically pleasing to someone with a highly imaginative personality.

Moving on to the interaction effect, the MANCOVA results for the interaction effect of shape x trait provided no significant results, suggesting that according to this data and the power of this sample, that personality is not a factor in preference for user interface skin shape.

In summary the results of the primary analysis did not provide any evidence to support H_3.

7.8 Limitations and Recommendations for Future Research

One of the main limitations of this experiment regards the shapes used, which were irregular polygons of varying numbers of turns from 5 to 40. These were derived from Martindale's polygons because they provided the opportunity to test the hypothesis in the context of relevant existing research. However, the Martindale polygons were designed as shapes in general to test arousal responses, and not applied as an interactive object such as a user interface. Typically user interfaces are rectangular and so the Martindale polygons produce user interfaces that are unfamiliar to the participant. Furthermore the designs of non-rectangular, irregularly shaped user interfaces do not use space optimally and therefore it is difficult to envisage an application where such randomly shaped polygons would be applied. Future research in this area could use experimental polygon shapes that better reflected real world user interface shapes, for example squares and rectangles. Additionally features of these polygons could be tested for aesthetic effect, for example rounded and square corners, where more acute angles might afford discomfort and soft curves afford comfort and ergonomics.

As for the previous hypothesis, H_2, the amount behavioural data collected was too small and therefore lacking sufficient statistical power for analysis. This was due to the same mechanism, a restriction on participants from receiving new skin packs until the appropriate personality and skin preference questionnaires had been completed. This was designed intentionally to prevent users from using the WebRadio player and contributing behavioural data without the associated personality and preference data, however the effect of this protocol meant a reduction in the number of participants that progressed through all phases of the experiment. The result of this meant complete sets of data for each participant but also had the consequence of reducing the overall amount of valid data. Future research ideally would address this limitation by increasing the number of participants and perhaps also by extending the length of the data collection period. Furthermore, future research in this area could concentrate on one particular design feature (e.g. colour) and then to provide fewer skin packs. By providing fewer skin pack conditions for participants to have to progress through would enable more participants in each condition. This would potentially provide a more useful dataset by using participants more optimally for statistical power. This research used 84 participants over a period of three months but after invalid data and participants were removed it resulted in 63 participants total over three months.

7.9 Conclusions

The results of the experiment did not support the experimental hypothesis, H_3. The MANCOVA analysis did not find trait a factor influencing user preference for skin shape, and furthermore that shape and trait together are not factors influencing user preference either. The analysis provided evidence to suggest that user preference for skins in general (i.e. unrelated to trait) is for fewer turned polygons but the data collected could not be linked to personality trait.

Two main conclusions can be drawn from these results. Firstly that shape is a factor in preference for user interface skins for certain personality traits, with preference declining as a function of the number of sides, and secondly, that no interaction effect could be observed between trait and shape in skin preference.

The next chapter investigates the fourth experimental hypothesis of this research. Whereas this and the previous chapter have looked at ways in which user interface

design features such as colour and shape might be linked to personality trait, H_4 looks at whether the meaning within an image can be linked to personality trait.

Chapter 8 – SKIN MEANING

8.1 Meaning: Overview

This chapter covers the investigation of the fourth experimental hypothesis relating to meaning within user interface skins. The second and third hypotheses investigated the effects of two user interface design factors, colour and shape that are operationalised by arousal. This fourth hypothesis (H_4) investigates the effect of meaning within a user interface skin, operationalised by the Similarity Attraction Hypothesis.

H_4: Users prefer user interface skins that exhibit a similar personality to their own

The chapter introduction sets the scene for this hypothesis by arguing that the Similarity Attraction Hypothesis may provide a means to provide a more personally relevant interaction, by matching user personality trait to a similar personality trait expressed graphically in a skin. The chapter then goes on to address how skins can feature personality traits and then how skins were created for use in this research. As part of the skin creation procedure a jury was used to verify the personality traits represented in the skins, and also to reduce a large pool of 50 skins down to a target figure of 7 skins for inclusion in the H_4 investigation. The details of the jury phases and methodology are then described, including a validation of the jurors to ensure that the sample of participants used were not significantly different from a normal global distribution. The next section addresses the final selection process to determine the skins that were on average most highly rated, and therefore verified by a jury to express a particular personality trait. Having discussed the creation and verification of the skins required to investigate H_4 the next section describes the method, results and conclusions for the investigation of H_4. This involved the collection and correlation of both user personality data and skin preference data. The chapter concludes with a summary of the H_4 investigation.

A visual guide to the chapter is shown below:

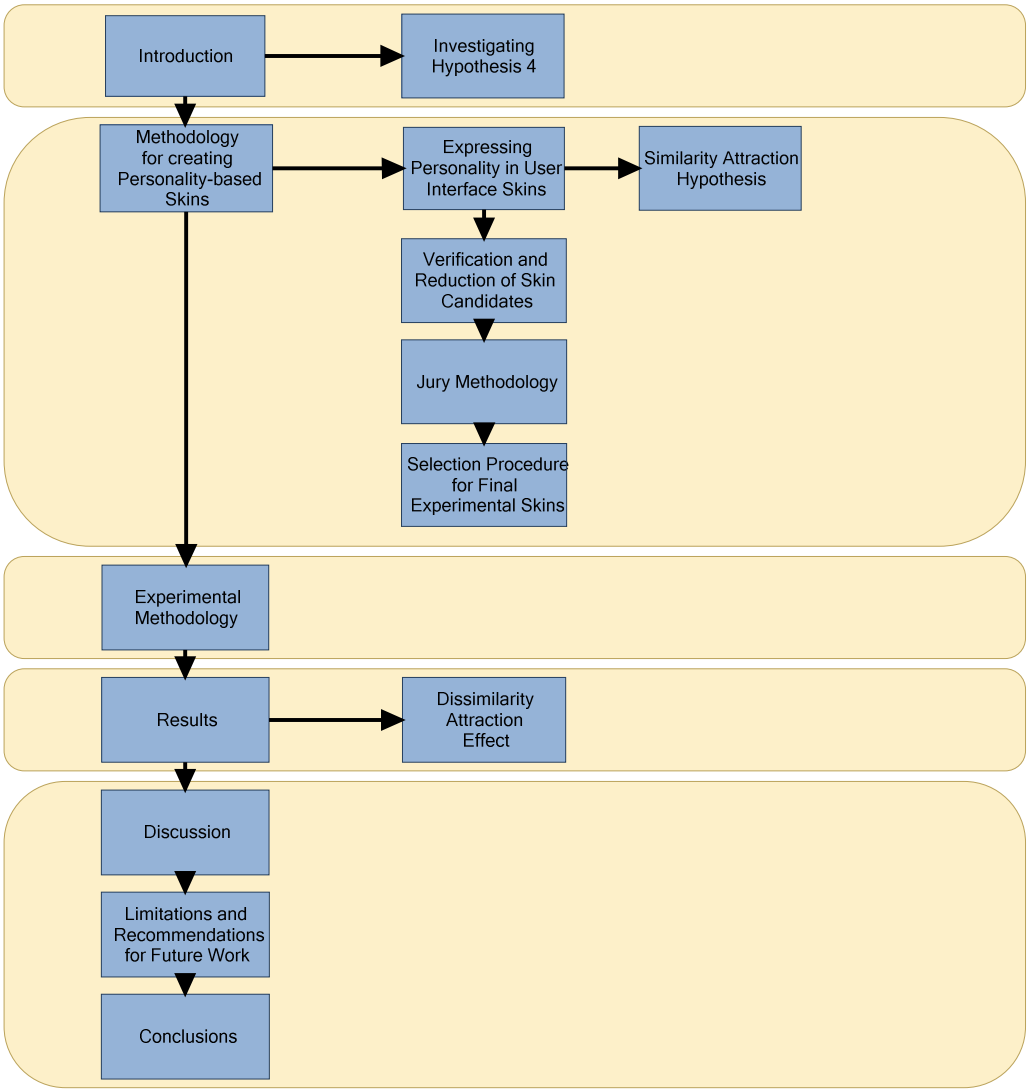


Figure 8-1 - Chapter overview

8.2 Introduction

In trying to create a more personally relevant interaction, the experimental approach has been to identify user interface design factors that may have greater appeal to certain personality traits, such as colour and shape. The purpose of H_4 is to investigate whether the perceived meaning within an image (e.g. personality trait), rather than aesthetic features (e.g. colour or shape) can be used as a user interface design factor.

Whereas colour and shape might not have an explicit meaning represented and contain no imagery or symbolic representation (e.g. purely flat background colours), representing meaning explicitly may provide a link to certain personality traits and facilitate design for personality trait. This is because the Similarity Attraction Hypothesis (Byrne and Nelson, 1965) suggests that there is a tendency to prefer people who exhibit similar personality traits to themselves. This idea has been extended by Nass et al. (1995) who suggest that computers can exhibit personality, and that humans behave socially towards computers. They tested whether the Similarity Attraction Hypothesis was supported for human computer interaction, as opposed to the interpersonal nature of the hypothesis proposed by Byrne and Nelson. They refer to the broader principle of humans acting interpersonally with computers as the "Computers Are Social Actors" (CASA) paradigm (Nass, Steuer and Tauber, 1994). When using a dominant and submissive scale they observed support for the Similarity Attraction Hypothesis, with dominant participants preferring to interact with a computer that exhibited a dominant personality and likewise submissive participants preferred submissive personalities.

If personality traits are expressed in pictures and can be measured it would provide an opportunity to determine whether the Similarity Attraction Hypothesis has any effect upon different user interface skins. If the Similarity Attraction Hypothesis is supported and it is found that participants prefer user interfaces that express similar personality traits to their own then it provides designers a means by which to segment, design accordingly, and ultimately provide a more personally appropriate interaction as compared to a single user interface.

A set of user interface skins which each expressed a personality trait were required, so that participant personality trait scores (IPIP-NEO) could be analysed against data collected from the skin preference questionnaire and interactive behaviour log files (e.g. frequency of skin selection, preferred skin). The next section address how Personality can be expressed in user interface skins.

8.3 Expressing Personality in User Interface Skins

This section argues that Personality can be expressed in images and other graphical representations (e.g animated characters). This is important because it provides an

opportunity to design a more personally relevant interaction by matching similar personality traits (i.e. that of the user and that of the skin).

Unlike colour and shape, which are both are quantifiable and broadly objective, the perception of meaning within an image is much more open to individual interpretation than colour or shape and therefore more subjectively perceived. For example, irrespective of individual differences, it is generally easier to gain agreement on whether an interface skin can be said to exhibit the colour red than the traits of extroversion. This creates a problem for research in this area because of the subjective determination of images by the experimenter. In order to address this issue, a jury approach was taken so that potential images could be rated by jurors on a personality scale in order to determine which images expressed which personality traits. By using a jury to rate the personality traits of each user interface skin, skins that were consistently marked extremely high or low on any of the personality trait dimensions could be retained and those that were marked consistently average (i.e. not appearing to express any particular personality trait strongly or weakly) could be discarded. This would serve to identify those images that were rated by participants as being strongly indicative of a particular personality trait or traits, and to confirm or reject the subjective selections made by the experimenter.

It is through the perception of personality in a static image that the Similarity Attraction Hypothesis is tested in H_4, to determine whether user interface skins that feature personality traits similar to the participant are preferred to the other available choices.

The expression and perception of personality has been observed in interaction with non living things, such as images, cartoons, games characters and websites. This demonstrates that an image can express personality. For example, Brown and Logan (2006) asked a participant jury to rate the personality types of characters of an animated television cartoon (*The Simpsons*). They found that jurors had a tendency to similarly rate the personality traits of the different characters, indicating a similarity in perception of expressed personality in a non living object.

Isbister (2005) describes how animated characters in computer games can be made to express personality. She suggests that extroversion and agreeableness are the most

legible traits, in that they are more quickly determined by the player than the other three traits of the five OCEAN factors, which can be determined over longer periods of time. She also states that exaggerating or highlighting personality traits in these characters can make them more appealing to the user. This suggests that not only can personality be expressed in computer generated objects and images, but that by doing so generated increased appeal to the user. As shall be seen in the following sections, user interface skins were created that expressed different personality traits within static images (photographs and illustrations) in order to test this hypothesis.

Brinkman and Fine (2005) conducted an exploratory study that found tentative evidence of the similarity attraction hypothesis in effect between personality trait and user interface skin preference. Although the study did not validate skins with a jury and as such are subjectively derived interpretation of the data, indications for similarity attraction were observed for a number of personality traits by correlations between trait and different types of skin (e.g. cartoon-like, scary, action theme).

8.3.1 Method for Creating User Interface Skins that Express Personality

The previous section has argued that personality can be expressed within a user interface skin. This section describes the method used to create the skins required to investigate H_4.

Initially a set of 50 royalty-free images were selected from a digital image library (Getty Images Inc.) by the experimenter using the IPIP-NEO questions (see **Table 34** below) as guiding factors by which to identify suitable images from the large digital image library. These selections would be then confirmed or denied by a jury, described in the next section.

id	Question
1	Worry about things.
2	Make friends easily.
3	Have a vivid imagination.
4	Trust others.
5	Complete tasks successfully.
6	Get angry easily.
7	Love large parties.
8	Believe in the importance of art.
9	Use others for my own ends.
10	Like to tidy up.
11	Often feel blue.
12	Take charge.
13	Experience my emotions intensely.
14	Love to help others.
15	Keep my promises.
16	Find it difficult to approach others.
17	Am always busy.
18	Prefer variety to routine.
19	Love a good fight.
20	Work hard.
21	Go on binges.
22	Love excitement.
23	Love to read challenging material.
24	Believe that I am better than others.
25	Am always prepared.
26	Panic easily.
27	Radiate joy.
28	Tend to vote for liberal political candidates.
29	Sympathize with the homeless.
30	Jump into things without thinking.
31	Fear for the worst.
32	Feel comfortable around people.
33	Enjoy wild flights of fantasy.
34	Believe that others have good intentions.
35	Excel in what I do.
36	Get irritated easily.
37	Talk to a lot of different people at parties.
38	See beauty in things that others might not notice.
39	Cheat to get ahead.
40	Often forget to put things back in their proper place.
41	Dislike myself.
42	Try to lead others.
43	Feel others' emotions.
44	Am concerned about others.
45	Tell the truth.
46	Am afraid to draw attention to myself.
47	Am always on the go.
48	Prefer to stick with things that I know.
49	Yell at people.
50	Do more than what's expected of me.
51	Rarely overindulge.
52	Seek adventure.
53	Avoid philosophical discussions.
54	Think highly of myself.
55	Carry out my plans.
56	Become overwhelmed by events.
57	Have a lot of fun.
58	Believe that there is no absolute right or wrong.
59	Feel sympathy for those who are worse off than myself.
60	Make rash decisions.
61	Am afraid of many things.
62	Avoid contacts with others.
63	Love to daydream.
64	Trust what people say.
65	Handle tasks smoothly.
66	Lose my temper.
67	Prefer to be alone.
68	Do not like poetry.
69	Take advantage of others.
70	Leave a mess in my room.
71	Am often down in the dumps.
72	Take control of things.
73	Rarely notice my emotional reactions.
74	Am indifferent to the feelings of others.
75	Break rules.
76	Only feel comfortable with friends.
77	Do a lot in my spare time.
78	Dislike changes.
79	Insult people.
80	Do just enough work to get by.
81	Easily resist temptations.
82	Enjoy being reckless.
83	Have difficulty understanding abstract ideas.
84	Have a high opinion of myself.
85	Waste my time.
86	Feel that I'm unable to deal with things.
87	Love life.
88	Tend to vote for conservative political candidates.
89	Am not interested in other people's problems.
90	Rush into things.
91	Get stressed out easily.
92	Keep others at a distance.
93	Like to get lost in thought.
94	Distrust people.
95	Know how to get things done.
96	Am not easily annoyed.
97	Avoid crowds.
98	Do not enjoy going to art museums.
99	Obstruct others' plans.
100	Leave my belongings around.
101	Feel comfortable with myself.
102	Wait for others to lead the way.
103	Don't understand people who get emotional.
104	Take no time for others.
105	Break my promises.
106	Am not bothered by difficult social situations.
107	Like to take it easy.
108	Am attached to conventional ways.
109	Get back at others.
110	Put little time and effort into my work.
111	Am able to control my cravings.
112	Act wild and crazy.
113	Am not interested in theoretical discussions.
114	Boast about my virtues.
115	Have difficulty starting tasks.
116	Remain calm under pressure.
117	Look at the bright side of life.
118	Believe that we should be tough on crime.
119	Try not to think about the needy.
120	Act without thinking.

Table 34 - IPIP-NEO (short version) questions, used to guide initial selection of image candidates for presentation to a subsequent jury (Source: Buchanan, T., Johnson, J. A., and Goldberg, L. R. (2005)

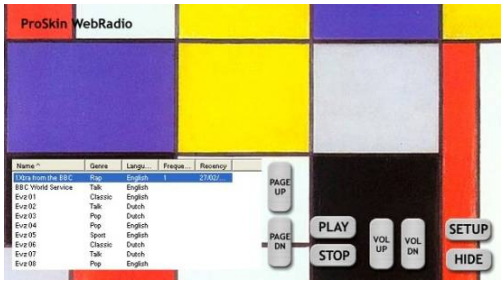
Potential image descriptions were created from the 120 IPIP-NEO questions. These were brief overviews describing the types of images to be selected from a digital image library, which would then comprise the initial set of images to be evaluated by the first phase of the jury testing. An example of this is represented below in **Table 35**, which shows the IPIP-NEO items and then the descriptions of potential images determined from them. The direction of the facet is indicated by a "+" for positive trait and "-" for negative in the "Sign" column. This table is used to illustrate the translation process from IPIP items into potential images for the initial image pool and the directions were subsequently either confirmed or rejected by juries and as such the potential image descriptions are important in this table, and the directions of lesser importance because they were the subjective first steps by the experimenter.

IPIP-NEO	Sign	Key	Facet	Item	Notes: potential image descriptions
14	+A3	A3	Altruism	Love to help others.	charity images
79	-A4	A4	Cooperation	Insult people.	fingers up to someone else
19	-A4	A4	Cooperation	Love a good fight.	rioting/demonstration
39	-A2	A2	Morality	Cheat to get ahead.	cheating in a card game - OBVIOUSLY cheating (e.g. 5 aces)
35	+C1	C1	Self-Efficacy	Excel in what I do.	winner/cup/podium - success image
100	-C2	C2	Orderliness	Leave my belongings around.	messy room and tidy room
75	-C3	C3	Dutifulness	Break rules.	thief/criminal and also police/judge/justice skin
57	+E6	E6	Cheerfulness	Have a lot of fun.	(people laughing at comedian)
67	-E2	E2	Gregariousness	Prefer to be alone.	desert island
107	-E4	E4	Activity Level	Like to take it easy.	maybe tandem image, rear rider feet up
87	+E6	E6	Cheerfulness	Love life.	smiley faces
1	+N1	N1	Anxiety	Worry about things.	OPPOSITE : use hammock, individual relaxing not worried
46	+N4	N4	Self-Consciousness	Am afraid to draw attention to myself.	stage view of busy/crowded auditorium - lots of faces looking at viewer
21	+N5	N5	Immoderation	Go on binges.	drunken people
116	-N6	N6	Vulnerability	Remain calm under pressure.	Schumacher - calm under pressure or similar... astronaut, air traffic control ?
63	+O1	O1	Imagination	Love to daydream.	working at pc, thought bubble of idyllic scene (daydreaming)
8	+O2	O2	Artistic Interests	Believe in the importance of art.	Mondrian and other classic art
68	-O2	O2	Artistic Interests	Do not like poetry.	poetry quotes
23	+O5	O5	Intellect	Love to read challenging material.	bookshelf with intellectual context
83	-O5	O5	Intellect	Have difficulty understanding abstract ideas.	large calculus/algebra

Table 35 - Translating IPIP-NEO items into potential images to be selected from a digital image library

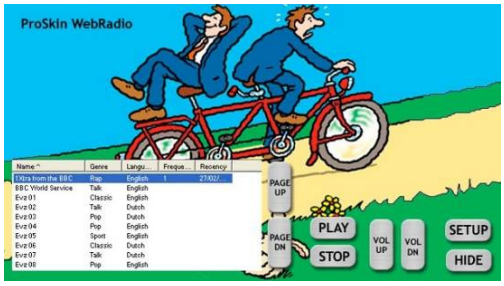
A number of the IPIP-NEO items are repeated during the 120 questions, so this resulted in an initial pool of 50 descriptions for which appropriate images were selected from a digital image library (Getty Images).

Figure 8-2 overleaf shows examples of images selected from the potential image descriptions for each of the five personality trait factors measured by IPIP-NEO.



Example 1: Openness to New Experience
Potential image description:
Mondrian and another classic art
IPIP-NEO item: Believe in the importance of art.
Facet: Artistic Interests

Example 2: Conscientiousness
Potential image description:
winner/cup/podium - success image
IPIP-NEO item: Excel in what I do.
Facet: Self-Efficacy



Example 3: Extraversion
Potential image description:
maybe tandem image, rear rider feet up
IPIP-NEO item: Like to take it easy.
Facet: Activity Level

Example 4: Agreeableness
Potential image description:
cheating in a card game – OBVIOUSLY cheating (e.g. 5 x aces)
IPIP-NEO item: Cheat to get ahead
Facet: Morality

Example 5: Neuroticism
Potential image description:
stage view of busy/crowded auditorium - lots of faces looking at viewer
IPIP-NEO item: Am afraid to draw attention to myself
Facet: Self Consciousness

Figure 8-2 - Examples for each OCEAN factor to illustrate image selection process

As a result of this process the initial pool of candidate user interface skins totalled 50, because the selected images had user interface elements added to them (e.g. buttons, radio station directory) which effectively turned them from static images to background images within user interface skins.

As shall be detailed in the next section, a process for verifying and reducing the candidate skins was required to reduce the total number of skins and to either confirm or deny the skins created by the experimenter.

8.3.2 Verification of and Reduction of User Interface Skin Candidates

This section describes how the initial pool of 50 user interface skin candidates were reduced in number and verified by a jury to confirm consensus perception of the personality traits expressed in the skins.

In the Brown and Logan (2006) research, a jury was used to establish expressed personality in *Simpsons* animated character images. A similar jury methodology has been employed by this research, firstly in order to verify the subjective image selection decisions taken by the experimenter and secondly to help refine and reduce the number of candidate skins for inclusion in the WebRadio skin distribution system.

In order to verify and reduce a pool of 50 candidate skins, four separate phases of jury were used. A target number of seven user interface skins was set for this hypothesis, to keep a similar number of user interface skin options available to the participant as for the other two skin-related conditions, colour and shape. In Hypothesis 2 there were seven different coloured skins in total available to the users and in Hypothesis 3 there were seven different shaped skins in total available to the users, so for consistency a similar number of personality user interface skins were sought. Four phases of jury were used so that at each phase candidate skins could be confirmed or removed from the selection pool.

In order to establish the perceived personality of the candidate skins, the jurors were asked to rate the perceived personality using a psychometric instrument. It was decided to use a shorter personality inventory than the IPIP-NEO because participants would be required to answer 120 questions for each of the 50 potential user interface skins, which results in a total of 6000 questions. It is not practical for one participant

to address so many questions because it becomes an unnecessarily long activity, and it could be argued that is ineffective use of participants because of the high attrition rate that could be expected with such a lengthy and repetitive task. Instead a shorter personality instrument was used, the Ten Item Personality Inventory (TIPI, Gosling et al. 2003). TIPI is a short ten item personality inventory that whilst is acknowledged by its authors as inferior to standard item questionnaires (such as IPIP-NEO) it also returns satisfactory correlations with IPIP-NEO and good test-retest reliability. The short number of questions required to establish personality traits makes the TIPI ideal for the jury studies because the largest number of questions asked of a juror is 500 (i.e. ten TIPI questions multiplied by fifty five potential skins).

By using a short personality inventory and multiple phases of jury assessment, the total number of skins could be reduced at each phase. The reduction in the number of skins meant that with each successive jury the participants would be required to answer fewer inventory questions.

The next section describes the phased jury approach adopted to address this issue.

8.4 Jury Phases

8.4.1 Overview

In this section the four phases of jury are described. The aim of the first pilot study was to test the methodology and to identify any immediately obvious unsuitable candidates. Skins were removed from the pool after the first pilot study and another pilot study run. This was because even though there were low numbers in the first pilot, the analysis suggested that there might be a gender effect confounding the results. Also, previous research (Brinkman and Fine, 2005) suggested that certain "humouristic" skins should be added to the candidate pool. These skins were added and the second pilot study completed. Having confirmed the methodology and the total number of candidate skins, two main jury studies were run to validate and reduce the skins to 18 in total prior to the final selection decision. The details of these phases are summarised below in Table 7.5, and the details of each phase in the following sections.

Phase in skin reduction process	Number of Skins	Number of Participants	Gender Distribution Male, Female
Pilot Study 1 (PS1)	50	6	4,2
Pilot Study 2 (PS2)	8	10	2,8
Main Jury Study (MJ1)	55	12	6,6
Main Jury Study (MJ2)	18	19	6,13
Final Skin Selection (MJ1+MJ2)	18	31	12,19

Table 36 - Summary of phases in the skin reduction process

8.4.2 Pilot and Jury Phases

This section describes the phases of the skin reduction task required to identify skins that the jury determined to exhibit a particular personality trait. The skins that were most strongly identified as having a particular personality trait would be used in the final selection process for the skins to be used in the web radio.

8.4.3 Pilot Study 1 (PS1)

An initial selection of 50 skins based on images obtained from an online digital library (Getty Images) were shown to 6 participants (4M, 2F) who were asked to rate the personality of the skin (using TIPI).

8.4.4 Pilot Study 2 (PS2)

Findings from Brinkman and Fine (2005) suggested that skins based upon gender and "cheerful" content should be also investigated. This was because firstly gender effects were observed in the Brinkman and Fine study and secondly because an analysis of PS1 suggested a participant preference for same gender skins. In addition to this a preference for cheerfully themed skins was highlighted, with correlations observed between extroversion and cheerful, humouristic and brightly coloured skins. Therefore an additional 8 skins were created that featured female specific images (as there were no female only skins in PS1 but there were male only skins) as well as cheerful images suggested by Brinkman and Fine (2005), specifically cartoon images of "*The Simpsons*" and "*SpongeBob Squarepants*" that were correlated with extroversion. These 8 skins were presented to a jury of 10 participants (2M, 8F) who were asked to rate the personality of each skin using the TIPI questionnaire.

8.4.5 Main Jury Study 1 (MJ1)

Images from PS1 and PS2 were combined into a pool of 58 images. In order to preserve an equal number of male-only and female-only images the three lowest scoring male-only skins were removed from the pool, leaving a total number of 55 images to be presented to the jury. The jury were asked to rate the personality of each skin using the TIPI questionnaire.

8.4.6 Main Jury Study 2 (MJ2)

The data from MJ1 was analysed and sorted by gender and mean TIPI rating for each of the five factors (OCEAN). The top three skins for each trait by mean were selected. Where two or more skins had the same mean, the skin with the lowest standard distribution was selected.

As a result 16 skins remained from the 55 in MJ1. An additional two skins were added in from MJ1 that had not been selected by highest mean because they had received consistently high means in PS1, PS2 and MJ1, bringing the total number of skins presented to the jury of MJ2 to 18.

The MJ2 jury (6M, 13F) were presented with the 18 skins and asked to rate the personality of each skin using the TIPI questionnaire.

8.4.7 Post Jury Final Skin Candidate Selection Procedure

As a result of the different phases of pilot and main jury studies a total of 58 skins were reduced to 18 skins. The results of MJ1 and MJ2 were combined, adding the data from MJ1 to MJ2 for the 18 skins in MJ2. This allowed for more data per skin to be analysed with the intention of reducing the effect of individual scoring outliers of jury members of the analysis (12M,19F) in the final reduction from 18 to 8.

A target number of 8 skins were set for inclusion within the WebRadio and so a final reduction strategy was required. A method was developed to transform the respondent's TIPI scores to make results further away from the mean appear greater, thereby highlighting extreme or high scoring results. A more detailed description of the problem and the method developed to solve it is presented below.

The TIPI questionnaire items are scored on a 7 point scale, with a score of 1 indicating low agreement and a score of 7 indicating high agreement. A score of 4 indicates neither agreement nor disagreement. Skins that are on average rated 4 (i.e. no preference) suggest that participants are indifferent to the trait expressed in the image and therefore such skins are of low value to this research because they are not perceived by the jury as being particularly expressive of any trait. As a result of this an analysis was designed that highlighted skins that were rated extremely high or low, as these skins would become candidates for the final skin pack to be included in the final skin pack used in the WebRadio. The details of the procedure now follow below.

8.4.8 Selection Procedure

The first step in the procedure was to place all TIPI items onto a scale in the same direction. TIPI features 10 questions, with two questions relating to each of the 5 traits. These questions address the two extremes of the trait, for example extroversion at one end and introversion at the other. Of the five traits, two have their directions reversed (agreeableness and emotional stability (the inverse of neuroticism)) so all data was transformed into a 14 point scale to make the direction of all traits uniform.

Having transformed the data onto a single direction scale for each trait, a function was applied to highlight the extreme scores at the low and the high ends of the TIPI scoring:

IF X < 8 THEN 8-X ELSE X-8

This function has the effect of transforming the mean personality trait data onto a new 7 point scale of extreme means. This was because skin candidates that were rated either high or low scoring on the TIPI would become candidates for the live experiment because these were the skins which on average were the most highly rated personality traits by the jurors (i.e. the extreme means). This is illustrated below in Figure 8-3.

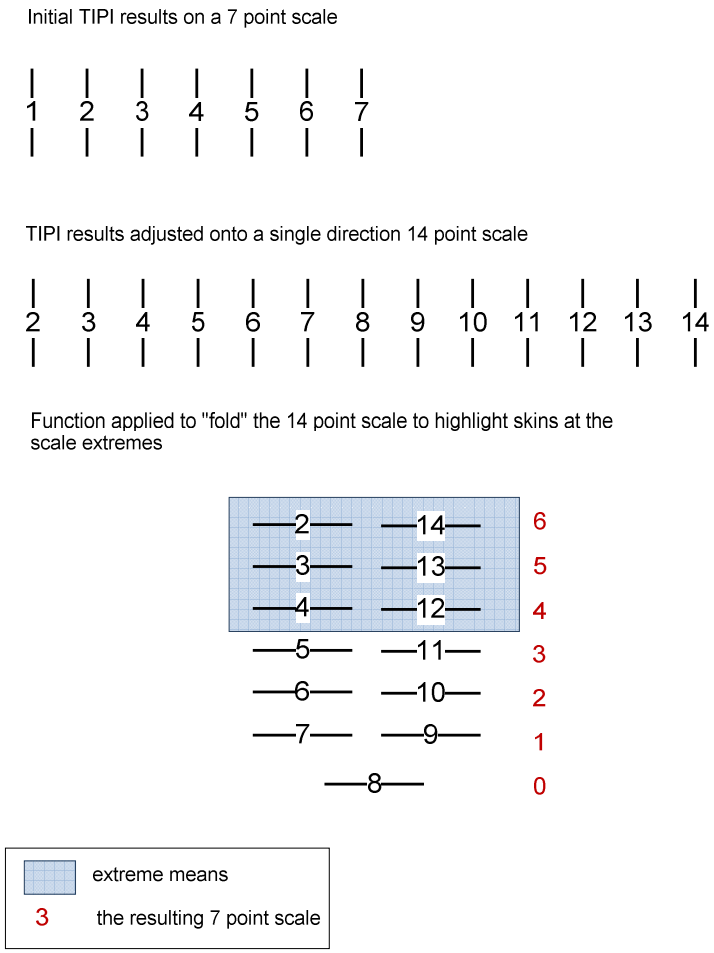


Figure 8-3 - Illustrating the transformation of data from TIPI to scale of extreme means

The results of the two main jury phases, MJ1 and MJ2, are summarised below in Table 37. The tables show the five OCEAN traits with the top three highest scoring extreme mean skins for both genders.

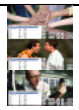
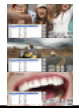
Agreeableness					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	15		11.7	2.6	3.7
	7		4.5	2.4	3.5
	49		4.8	2.3	3.2
Female	34		2.5	3.3	3.7
	1		12.5	2.9	3.2
	6		4.5	2.2	3.0
Conscientiousness					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	41		4.0	2.4	4.0
	37		11.8	2.3	3.8
	1		10.3	3.0	2.3
Female	41		4.0	3.0	3.3
	1		11.0	2.6	3.0
	37		11.5	2.3	2.8
Extraversion					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	28		11.8	3.0	3.8
	33		11.5	2.9	3.5
	3		10.7	2.8	2.7
Female	28		13.5	2.6	4.2
	19		12.5	2.3	4.0
	29		11.5	2.3	3.7
Emotional Stability (inverse of Neuroticism)					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	49		4.3	2.7	3.7
	34		4.5	2.8	3.5
	12		11.3	2.2	3.3
Female	1		12.0	2.7	3.3
	16		5.0	2.3	3.0
	49		5.0	2.6	2.8
Openness to New Experience					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	1		9.8	2.8	2.7
	33		9.8	1.8	1.8
	4		9.3	1.2	1.3
Female	5		11.0	2.0	2.3
	16		10.0	1.7	2.2
	35		9.0	2.1	1.5

Table 37 - Summary of skin reduction from 55 to 18 in MJ1 (n=12 (6M, 6F))

Agreeableness					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	8		11.5	2.1	3.5
	6		1.8	2.2	3.3
	18		4.8	1.9	3.2
Female	9		4.6	2.5	3.4
	14		4.7	2.5	3.3
	1		11.2	2.1	3.2
Conscientiousness					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	17		4.0	2.3	4.0
	16		11.3	2.3	3.3
	1		10.8	2.3	1.8
Female	1		11.4	2.6	3.4
	17		4.7	3.2	3.3
	16		11.1	1.7	3.1
Extroversion					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	11		11.8	2.6	3.8
	3		11.6	2.0	3.6
	10		11.4	2.2	3.4
Female	11		12.0	2.0	4.0
	3		11.5	2.4	3.5
	12		11.2	2.0	2.4
Emotional Stability (inverse of Neuroticism)					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	18		4.6	2.4	3.4
	14		4.9	2.3	3.1
	1		10.8	3.3	2.8
Female	9		4.9	1.9	3.2
	1		10.9	2.0	2.9
	7		10.9	2.3	2.9
Openness To New Experience					
Gender	Skin Number	Image	TIPI		Extreme Mean
			Mean	SD	
Male	10		9.8	1.9	1.8
	1		9.5	1.5	1.5
	13		9.2	1.5	1.2
Female	9		9.6	1.6	1.6
	4		9.1	1.7	1.1
	10		9.1	1.9	1.1

Table 38 - Summary of skin reduction from 18 to 8 in Final Selection (n=31 (12M, 19F))

The top skins from each of the five personality traits were selected for each gender, except in the case of an equal score when the lowest standard deviation was used to determine which was selected. For the female neuroticism skin, skin 7 was selected instead of the highest extreme mean for that trait because the top two skins for that trait were already included in the final skins. This represents a different selection strategy to extroversion where one skin was selected to represent both male and female extroversion. The reasoning for this change is because the extrovert skins show much higher extreme means than for the neurotic skins and therefore a more likely to be effective in representing a particular trait to the jury. Therefore for female neuroticism it was decided to use the next highest extreme mean from a neurotic skin that was not already represented in the final skin selection. This produced 10 skins for the final skin pack. However, the scores for the openness to new experience trait scores were all very low for both male and female results (less than 3 on the extreme mean). This suggested that the images in the openness skins were not perceived as having any particular openness personality traits and therefore not valid for inclusion in the final skin pack. This left 8 skins that comprised the skins to test perceived personality and Hypothesis 4, illustrated below in Figure 8-4. The final figure was 6[6] because the extroversion trait was represented by the same skin for both genders.

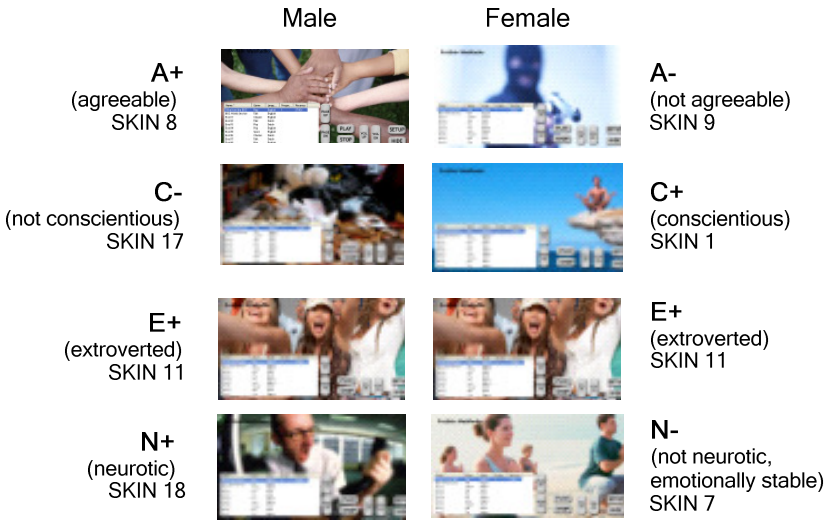


Figure 8-4 - Final skins for inclusion to test Hypothesis 4

[6] The skin for male agreeableness was mistakenly omitted from the skin pack uploaded to the ProSkin server and as a result participants did not receive the skin. This skin was removed from the analysis leaving a total of 7 different personality trait skins presented to participants during the experiment.

8.4.9 Validation of the Jury

The jury data (defined as the data from PS1, PS2, MJ1 and MJ2, n=43) was checked for internal reliability and compared with the normative data used in the TIPI questionnaires. This was to ensure two things - firstly that juror responses were internally consistent across the questionnaire and secondly to ensure that the personality profile of this experimental jury was not significantly different from global norms.

In order to test the internal reliability Cronbach's Alpha was used. Results showed high internal reliability for each of agreeableness, conscientiousness, extroversion and neuroticism dimensions and average reliability for openness. This confirmed the decision to remove the openness skins on the basis of low extreme average scores (i.e. indifferent or average perception of all traits). The results are summarised below in Table 39:

	Total alpha	Male alpha	Female alpha
O	0.606	0.664	0.65
C	0.955	0.879	0.929
E	0.938	0.819	0.911
A	0.967	0.907	0.951
N	0.953	0.843	0.938

Table 39 - Cronbach's Alpha data for all jurors, showing high reliability on the ratings of the A, C, E and N dimensions.

These results demonstrate that the participants in this study were able to perceive the same personality trait in a user interface skin and replicates the findings of Brown and Logan.

In order to test the population validity of the jurors a one-sample t-test was performed on personality trait scores of the jurors themselves to compare the means with the normative data means used to construct the TIPI questionnaire (Gosling et al., 2003). This would determine whether personality trait means of the jurors used in this research were significantly different from those used to establish the TIPI instrument.

The results showed that for extroversion, conscientiousness and neuroticism that there was no significant difference.

The results showed no significant difference between the personality traits, with the exception of agreeableness and openness to new experience. This indicates that the jury used in this investigation were not representative enough of a mean general population for these traits. These findings mean that the results should treated with some caution for these traits because the experiment jury were less agreeable and less open to new experience than a general population on average. Furthermore, this also likely accounts for the fact that it was not possible to create an openness skin. The results are summarised in Table 40:

	E	A	C	ES (N)	O
All					
Normative Mean	4.44	5.23	5.4	4.83	5.38
Jury Mean	4.7	4.63	5.4	4.65	4.46
Jury SD	1.36	0.927	1.1	1.32	0.571
t	1.2	-3.52	<0.001	-0.747	-8.7
sig (two-tailed)	0.24	0.001	1	0.461	<0.001
df	29	29	29	29	29
Male					
Normative Mean	4.25	5.06	5.19	5.13	5.34
Jury Mean	4.04	4.4	5.13	4.59	4.54
Jury SD	1.49	0.917	1.18	1.81	0.471
t	-0.455	-2.3	-0.15	-0.986	-5.58
sig (two-tailed)	0.659	0.04	0.884	0.348	<0.001
df	10	10	10	10	10
Female					
Normative Mean	4.54	5.32	5.51	4.66	5.4
Jury Mean	5.07	4.76	5.55	4.68	4.42
Jury SD	1.16	0.933	1.05	0.988	0.629
t	2	-2.6	0.177	0.107	-6.779
sig (two-tailed)	0.6	0.18	0.862	0.916	<0.001
df	18	18	18	18	18

ES represents Emotional Stability (TIPI) which is the inverse of Neuroticism (IPIP-NEO)

Table 40 - Summary of one-sample t-test for both genders, male only and female only comparing personality trait data of the experimental jury with the normative TIPI personality trait data (Gosling, 2003)

The previous content of this chapter has addressed the creation of a set of skins that represent particular personality traits. The next sections describe the investigation of H_4 using the personality trait skins in the WebRadio application.

8.5 Method

Sixty four participants downloaded and installed the ProSkin WebRadio on their own personal computers in their own context of use. They completed electronic versions of both a personality inventory (IPIP-NEO) and skin preference questionnaire and the results transmitted via the Internet and stored in the ProSkin database. Once the questionnaires were completed a package of the seven final H_4 skins were uploaded to each client (two skins for Openness To New Experience were removed because the jury did not rate them strongly as representing a particular personality trait and one skin for male agreeableness was mistakenly omitted) . These meaning based user interface skins were made available to the user and could be selected by the user by clicking on the 'setup' button. All user requested skin changes were recorded to log file and stored in the database.

8.6 Analysis Methodology

This section provides a description of the analysis methodology used to investigate H_4.

The experimental hypothesis investigated in this chapter states that users will prefer user interface skins that exhibit similar personality traits to their own. As a result of this, seven experimental user personality trait interface skins were developed using juries. In order to find evidence to support H_4 users would need to indicate a preference for personality trait skins that were determined by the juries to be of similar personality trait to their own, in accordance with the Similarity Attraction Hypothesis. Specific Pearson's correlations were performed between the personality trait of the participants (e.g. extroversion) and the corresponding skin selected by jury to represent that particular trait. Significant correlations at 5% alpha that were in the predicted direction (e.g. neurotic/not neurotic) would be considered as evidence to support the experimental hypothesis.

In the investigation of H_2 and H_3 the correlations are performed as post hoc analyses whereas for this hypothesis they are being performed as a priori analyses. This is because the experiment has been designed from the outset to investigate the Similarity Attraction Hypothesis, which suggests that people prefer other people with similar personality traits to their own. The a priori hypothesis, that users will prefer user interface skins that exhibit similar personality traits to their own, is tested by specifically directed correlations. These correlations are directed by the personality traits of the user interface skins, so that in order to support H_4 significant correlations between the specific user interface skin trait and user preference would need to be found. For example, the extrovert skin should be preferred significantly more by participants with high extroversion scores. Other traits may also show significance for a specific user interface skin although the jury did not consider these skins to exhibit such traits and therefore are undirected correlations and do not test the experimental hypothesis.

As noted in the previous two chapters, there were significantly more males than females that participated in the experiment (54 males, 9 females). As a result of this imbalance only the male participant data has been analysed. However, because the female skins were also seen by male participants and data collected, these skins were added to the male skins for analysis purposes using the male response data. This meant using the female skins in the male analysis, although these two analyses have been kept separate. The original set of male skins are referred to as Primary Analysis I and the additional female skins with male data are referred to as Primary Analysis II. Primary Analysis II is conducted as an a priori analysis because the jury had determined the personality traits of those skins prior to the commencement of the experiment.

No behavioural data was used in the investigation of H_4. This is because there was not sufficient data collected to be able to perform the correlation analyses.

From this point in the analysis the TIPI scores are referred to on a 1-7 scale as opposed to the 2-14 scale for ease of explanation. This is because the directions of each trait do not require any change in direction by the reader and as such make for easier reading.

8.7 Results and Observations

		Primary Analysis I			Primary Analysis II		
Skin Number		17	11	18	9	1	7
Factor		C	E	ES	A	C	ES
							
Jury TIPI Score	E	4.33	5.92	5.04	4.58	4.50	4.08
	A	3.13	4.79	2.42	2.67	5.13	5.04
	C	2.00	3.63	3.54	3.83	5.38	5.17
	ES	3.33	4.29	2.29	3.33	5.38	5.29
	O	4.29	4.25	4.04	3.75	4.75	4.17
Participant Personality				**Correlations**			
Gregariousness (E)		-0.287*					
		0.035					
Excitement Seeking (E)			**0.323***				
			0.017				
Agreeableness (A)		-0.296*					
		0.030					
Cooperation (A)			-0.274*				
			0.045				
Altruism (A)							**0.390****
							0.004
Achievement Striving (C)		**-0.324***					
		0.017					
Cautiousness (C)						*-0.310**	*-0.342**
						0.022	0.011
Neuroticism (N)						*0.435***	*0.357***
						0.001	0.008
Anxiety (N)						*0.472***	*0.273***
						0.001	0.046
Anger (N)						*0.282**	*0.443***
						0.039	0.001
Depression (N)						*0.336**	
						0.013	
Vulnerability (N)						*0.324**	*0.311**
						0.017	0.022
Emotionality (O)							0.275*
							0.044

* indicates significant at p=<0.05 (two-tailed), ** indicates significant at p=<0.01% (two-tailed)
correlations are listed r value first with sig (2-tailed) underneath
Jury TIPI Score uses Emotional Stability, Correlations use Neuroticism
bold values indicate support similarity attraction hypothesis
italicised values indicate support dissimilarity attraction hypothesis
underlined values in Jury TIPI section indicate extreme trait score
extreme trait score defined as scores lying outside of the 3-5 range

Table 41 - Significant Pearson Correlations results, showing the personality trait scores as determined by the jury for the meaning skin pack with associated correlation (p-value) between participant personality trait and their skin preference. Male participants.

The output from the correlations showed mixed results with evidence to support the hypothesis and also evidence contrary to the experimental hypothesis. To briefly summarise the results of each trait investigated, extroversion and agreeableness were found to support the experimental hypothesis of similarity attraction ; neuroticism was found to support the opposite of similarity attraction (referred to in this research as dissimilarity attraction), and conscientiousness provided conflicting results, showing both similarity and dissimilarity attraction. These results are now discussed in greater detail. The data will be interpreted according to the experimental hypothesis, testing to see whether the similarity hypothesis is confirmed or not, and whether any dissimilarity effect is observed for each trait.

According to the Similarity Attraction Hypothesis, one would expect to find correlations between the personality trait exhibited by the participant and the corresponding personality trait exhibited by the user interface skin. For example, highly extrovert individuals should prefer the highly extrovert skin (a positive correlation) and low extrovert individuals should prefer the low extrovert skin (a negative correlation).

Extroversion was represented by an image of energetic women and was significantly positively correlated with the excitement seeking facet of Extroversion (r=0.323, sig (2-tailed) 0.017). The result provides support for the experimental hypothesis and appears to demonstrate similarity attraction. This also provides support for Reeves and Nass (1996) who found support for the Similarity Attraction Hypothesis when examining the dominant or submissive interaction styles. They refer to the dominance-submission dimension as synonymous with extraversion (1996, p.75):

"…we are at opposite ends of a single dimension of personality called extraversion or dominance/submissiveness."

Similarly, Matthews, Deary and Whiteman (2003) list the traits associated with the three dimensions of Eysenck's model of personality. For extroversion the adjective "dominant" is listed, providing additional evidence to support the link between dominance and extroversion.

Neuroticism was represented by three skins. Skin 18 was rated as highly neurotic by the jury and depicted a stressed man on the telephone. Skin 7 is an image of women

meditating on a beach, an image selected by the jury as representing a highly emotionally stable/non-neurotic personality trait. Skin 1 is an image of a man meditating on a rock and is seen as the jury as being non-neurotic. The analysis revealed a positive correlations between Skin 7 and the main factor of neuroticism (r=0.357, sig (2-tailed) 0.008) and the facets Anxiety (r=0.273, sig (2-tailed) 0.046) and Anger (r=0.443, sig (2-tailed) 0.001), and positive correlations between Skin 1 and the main factor of neuroticism (r=0.435m sig (2-tailed) 0.001) and the facets anxiety (r=0.472, sig (2-tailed) 0.001), anger (r=0.282, sig (2-tailed) 0.039, depression (r=0.336, sig (2-tailed) 0.013) and vulnerability (r=0.342, sig 0.017).

These results suggest that individuals that score highly on these traits prefer a non-neurotic skin, which is opposite to the expected similarity attraction and is considered dissimilarity attraction for the purposes of this research. One interpretation of these findings is that for what might be commonly considered a "negative" trait, such as neuroticism and its facets, rather than individuals exhibiting a tendency to prefer similar personalities, they exhibit a tendency to prefer opposite personality traits that might reduce any negative effects that highly neurotic individuals experience (i.e. aspirational, or equilibrium seeking behaviour). In this instance the calming imagery of peaceful women meditating in a serene environment is likely more attractive to a highly neurotic individual than imagery relating to high neuroticism that would exemplify similarity attraction behaviour. It also may help to explain why there were no significant correlations observed for the neurotic skin (skin 18) and why correlations were observed for the non-neurotic skin (skin 7), because a neurotic person would not be attracted to a neurotic personality. There is insufficient data to determine whether a non-neurotic person would be attracted to a neurotic personality.

Skin 1 is designated an agreeable skin, however, it is also a non-neurotic skin because it was rated by the jury to be extremely non-neurotic (see Table 41). The correlations observed between skin 1 as a non-neurotic skin were in the same direction as the correlations for skin 7, which adds further evidence to support the theory that there is a dissimilarity effect happening. From this data there is a strong case for dissimilarity attraction for the neuroticism trait.

No significant correlations were observed for Skin 18. There are a number of reasons as to why significant correlations were not observed for Skin 18. It might have been

the case that there was insufficient power in the study and a higher number of participants would have given the opportunity to find a small significant correlation. Another possible reason is that it was the only low emotional stability skin of the three, with the other two representing high emotional stability, and low emotional stability is likely a more desirable personality trait to both low and high neurotics. This is because on the basis of these results and in the context of similarity/dissimilarity attraction, it seems that a low emotional stability skin would appeal to limited participants because it doesn't appear to provide any obvious aspirational or equilibrium-seeking appeal. It might be that similarity attraction is only observed when the image could be considered to be positive, or not negative.

The agreeableness skin was depicted by an image of a masked gunman (skin 9), determined by the jury to represent non-agreeableness. No significant correlations were observed. However, the non-neurotic skin (skin 7) was also rated as being highly agreeable by the jury, and a significant correlation was observed between the altruism facet of agreeableness (r=0.390, sig (2-tailed). Although this provides evidence to support the experimental hypothesis, because it is the altruism facet it is possible that this finding does not represent similarity attraction to the user interface skin but instead represents the participation of highly agreeable participants and as such may represent an experimental confound. However, if it was a strong effect it would likely have produced multiple correlations for the altruism facet across more skins than observed and not skin 9 in particular.

For conscientiousness there were two skins determined by the jury, one that represented low conscientiousness depicting a messy room (skin 17) and one that represented high conscientiousness depicting a meditating man (skin 1). A significant negative correlation was observed between the messy room skin and the achievement striving facet of conscientiousness and (r=-0.324, sig (2-tailed) 0.017), suggesting that participants with low conscientious personality trait scores had a preference for the non-conscientious user interface skin. This finding supports the experimental hypothesis. However, a significant negative correlation was observed between the highly conscientious skin (skin 1) and the cautiousness facet of conscientiousness (r=-0.310, sig (2-tailed) 0.022). This result indicates dissimilarity attraction and unlike any of the other three traits investigated provides conflicting

evidence within the conscientiousness trait. This is because significant correlations in the opposite directions have been found for the same trait, indicating both similarity and dissimilarity attraction. The results for this trait do not support the similarity or dissimilarity view. It was likely caused by another factor because for the non-conscientious skin (skin 17) it makes sense that a conscientious person would not be attracted to it, but the conscientious skin was disliked by people who were highly cautious. The analysis of the neuroticism correlations suggested that perhaps negatively perceived traits such as neuroticism would be attracted to emotionally stable personalities but yet this same kind of dissimilarity attraction does not appear to occur for conscientiousness. If a similarity attraction effect was happening for cautious individuals then one would expect a positive correlation rather than the negative correlation observed. A scatter plot was created to check whether there were any outlying data that might have affected the results but no clustering or prominent outliers could be seen. The original TIPI questionnaire adjectives were checked to see if they could be ambiguous or somehow relate to the image used in the experiment skin, and no ambiguity could be found to explain these results. Additional discussion regarding these contrary findings is found in the discussion section below.

In testing the hypothesis that users will prefer a user interface skin that exhibits a personality trait similar to their own, these mixed results provide evidence to reject the Similarity Attraction Hypothesis as applied to personality in general across all five factors. The experimental hypothesis, H_4, was supported by the results for extroversion. However, strong evidence for a dissimilarity attraction effect for neuroticism presented here indicates that the experimental hypothesis can be reconsidered in terms of trait, with similarity and dissimilarity attraction a function of personality trait.

The similarity/dissimilarity attraction findings are summarised in Table 42 :

Personality Trait	Predicted	Observed	Facet	Similarity Attraction?	Dissimilarity Attraction?
Agreeableness	###	###			
Conscientiousness	C-	C-	achievement striving	yes (1)	
	C-	C+	cautiousness		yes (2)
Extroversion	E+	E+	excitement seeking	yes (1)	
Neuroticism	N-	N+	**Neuroticism**		yes (2)
			anxiety		yes (2)
			anger		yes (2)
			depression		yes (1)
			vulnerability		yes (2)
Openness To New Experience	###	###			

indicates no skin
number of correlations in brackets
bold indicates main trait

Table 42 - Summary of Similarity/Dissimilarity Attraction Hypothesis Testing Results

8.8 Discussion

The results from the analysis were mixed, providing both similarity and dissimilarity attraction results. Whilst the experimental hypothesis was supported from the extroversion results it is clear from the strong evidence provided for neuroticism that there is a dissimilarity attraction effect. The third factor investigated, conscientiousness, provided evidence to support the experimental hypothesis and evidence to support a dissimilarity hypothesis. These contradictory results for conscientiousness may provide further insights as to which conditions are required for either similarity or dissimilarity attraction to occur.

An explanation for these results is that conscientiousness as a construct does not evoke similarity attraction across the facets in uniform directions, as it does for neuroticism and possibly for extroversion. Another explanation is that conscientiousness may work on a per facet basis, and within each facet there are different similarity effects according to the level of a particular facet (e.g. high cautiousness compared with low cautiousness when considering similarity or dissimilarity attraction to a particular expressed trait). Table 43 summarises the similarity and dissimilarity effects observed in this experiment, and also provides two cases from the Reeves and Nass dominance-submissive study on the extroversion trait to try to understand the nature of similarity and dissimilarity attraction.

	User interface skin position on trait dimension	
Trait	**Low**	**High**
Open to New Experience		
Conscientiousness	Similarity (1 skin)	Dissimilarity (2 skins)
Extroversion	Similarity (1 R&N UI)	Similarity (1 skin + 1 R&N UI)
Agreeableness		
Neuroticism	Dissimilarity (2 skins)	
R&N indicates user interface from Reeves and Nass (1996) study		

Table 43 - Summary of Similarity and Dissimilarity Effect as a function of trait and position on trait dimension

8.9 Limitations and Recommendations for Future Research

One of the main limitations to this experiment has been the gender imbalance and the analysis of only male datasets, which has meant that any conclusions drawn are based solely upon male participants. Future research in this area with balanced participant datasets would be able to better determine any gender related differences regarding similarity or dissimilarity attraction.

Another limitation of this study is the fact that some user interface skins feature a single person and others feature multiple. In terms of similarity attraction, when there are multiple people represented in a skin, it might offer the participant multiple targets for any empathy or abstraction of trait from. For example, in the skin for the facet of extroversion, featured on page 152, there are two people represented. One is pedalling hard uphill and the other is sitting with his feet up. Clearly both people have different activity levels at the seeming extremes of the facet, the high activity individual pedalling and the low activity individual not actively participating. In terms of representing a trait, it seems like having either a single person or group expressing the intended trait is less ambiguous to the viewer and may produce more consistent results because only one direction of the trait is explicitly offered to the viewer. This skin in particular was not selected by the juries for inclusion in the final set of experiment skins. Future research should ensure uniform direction of expressed traits in user interface skins presented to jurors and participants to minimise this confound.

This experiment has investigated whether the similarity attraction effect can be observed for three of the Big Five personality traits (OCEAN). There is some support

for extroversion and the Similarity Attraction Hypothesis when applied to personality traits represented in user interface skins, but there is limited research available that addresses the other four traits. Future research could focus on one particular personality trait to determine whether similarity or dissimilarity attraction occurs, and under what conditions. Other traits could be investigated specifically with the intention of investigating this conditionality.

A method has been developed here for establishing perceived personality trait in user interface skins by using a jury approach. This provides a more objective method than the previously adopted subjective categorisation of user interface skins in Fine and Brinkman (2005) that was dependent upon the experimenter's perception and interpretation of the skins as opposed to a multi phase jury. Future work could refine the jury methodology, including number of phases and selection criteria, to provide a more efficient method for establishing perceived personality trait in user interface skins.

8.10 Conclusions

The overall conclusions drawn from this experiment are as follows:

- People appear to consistently perceive personality expressed in user interface skins.
- There appears to be a relationship between participant personality and preference for user interface skin that can be explained by the perceived personality.
- The experimental hypothesis, H_4, that "users will prefer a user interface that exhibit a similar Personality to their own", does not appear to hold true for all personality traits.

The results of the analyses revealed both similarity and dissimilarity attraction. Whilst there is some evidence to support the experimental hypothesis there is also evidence to suggest the inverse of the experimental hypothesis, that certain personality traits prefer user interface skins that do not exhibit personality traits like their own, or specifically that certain personality traits prefer user interface skins that exhibit personality traits desirable to them. Furthermore it appears that user

preference for user interface varies as a function of personality trait and position on the trait scale, so the experimental hypothesis can be revised to make it specific to personality trait, so a revised hypothesis is suggested for future work in this area:

H_4 revised *Users preference for user interface varies as a function of personality trait and position on the personality trait scale*

Further research is required to be able to determine whether the dissimilarity attraction effects observed here are repeatable and externally valid, and then to try to determine the conditions regulating similarity or dissimilarity attraction for each of the five factors. Additionally, the jury-based method for establishing perceived personality trait in user interface skins can be developed to provide greater efficiency, potentially fewer jury phases and possibly different skin selection criteria to the extreme means used here. Using a jury-based selection process may provide designers a means to validate their designs for personality. For example, if through a process of log file psychometrics (described in Chapter 4) or otherwise (such as traditional psychometric instruments), designers are able to know the personality profile of their intended users, then an ability to produce a user interface skin that consistently exhibits the intended personality trait is required. A jury-based methodology, such as the one proposed here, gives designers a means to validate their own designs to ensure the intended personality trait is the perceived personality trait.

Having investigated all four experimental hypotheses the next chapter provides a conclusions chapter, summarising the results of the hypothesis testing and any insights learned.

Chapter 9 – CONCLUSIONS AND FINAL REMARKS

9.1 Conclusions

This chapter draws together the conclusions from the testing of the four experimental hypotheses and serves to summarise and conclude this thesis. The chapter starts out with a recapitulation to address the intentions and conclusions drawn from the research and then features four sections each summarising the conclusions from the experimental hypotheses. The next section addresses the overall experimental hypothesis, that a more personally relevant interaction can be produced by using personalised user interface skins. The section afterwards provides a reformulation of the experimental hypotheses by way of recommendations for future work. The following section presents the contributions arising from this work. The final remarks section concludes the thesis. A visual guide to the chapter can be found in Figure 9-1.

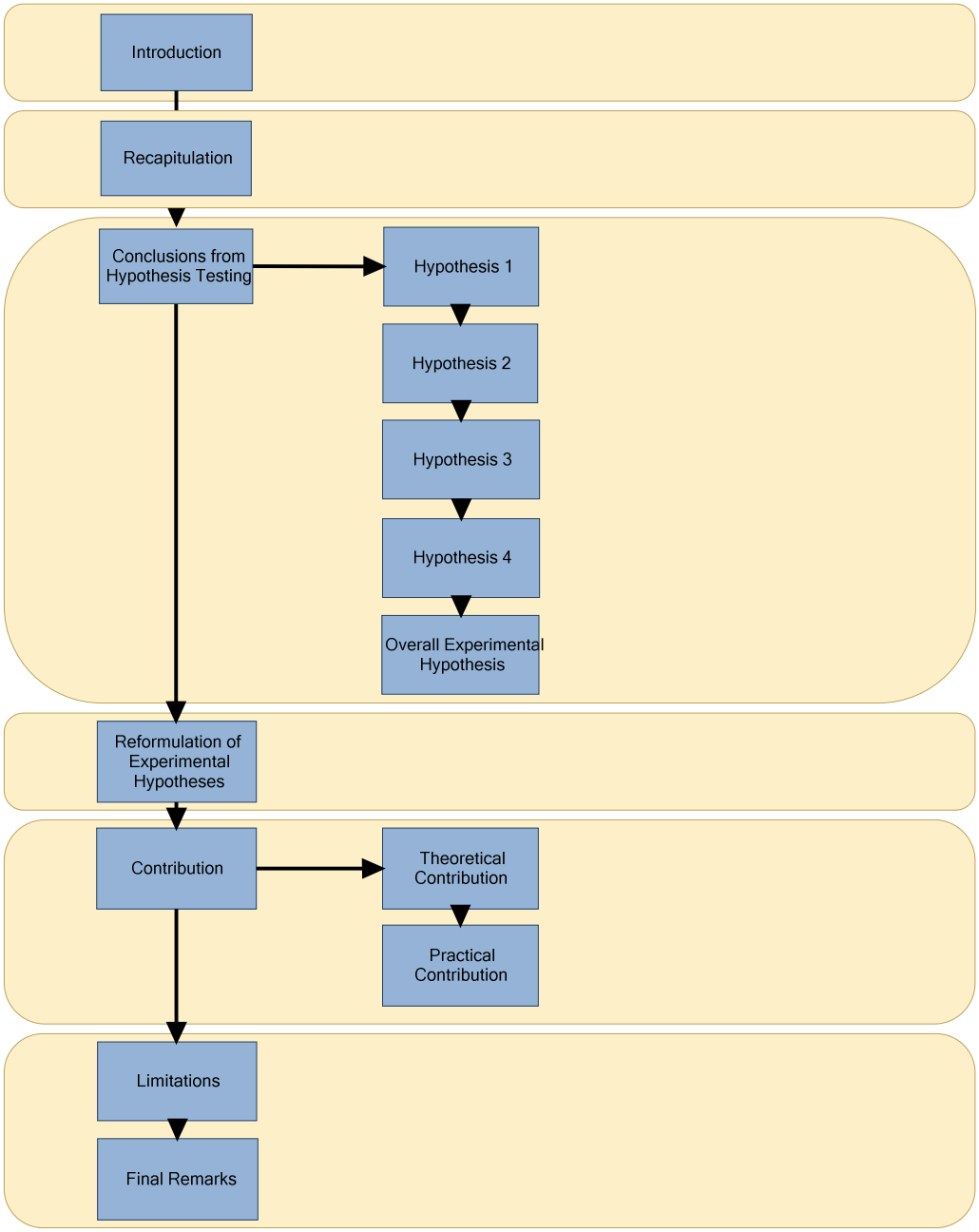


Figure 9-1 - Visual guide to Chapter 8

9.2 Recapitulation

At the outset of this thesis the motivations for this work were stated: to provide a more personally-relevant interaction using user interface skins than the one offered by a single user interface approach designed for the average intended user, and in so doing inform users and designers as to how user interface skins can be used to contribute towards a more personally appealing user experience. As stated in Chapter

1, one of the issues is that a number of software applications today support user interface skins yet there is relatively little research to inform users and designers how user interface skins based on personality can be used to enhance the user experience. Through the empirical testing of four hypotheses a method for personalising interaction using user interface skins has been investigated. The method involves the measurement of personality using interactive behaviours and then providing user interface skins designed for a particular personality profile. Links between personality trait and both colour and meaning in skins have been demonstrated, suggesting that these two interface design elements can be modified by designers to appeal to specific personality traits. By providing personality profiled skins to the appropriate users a more personal interaction is offered in comparison to a single user interface approach.

9.3 Conclusions Drawn From Hypothesis Testing

9.3.1 H_1: Personality can be extracted from log files

Weak evidence to support the experimental hypothesis was found, with the maximum variance in personality trait accounted for by an interactive behaviour at 49%. The data suggests that there are factors in addition to the studied behavioural measures that can explain personality trait, so whilst personality can be extracted from log files according to this data, it is not possible to measure personality from log files with the same level of accuracy as the IPIP-NEO according to this data.

Future work could seek to establish additional metrics for inclusion in the predictive model to try to improve trait variance accounted for by personality trait beyond the 49% observed here. These metrics could be extensions to the sets created for this experiment, for example keyboard button presses or relative volume level, and they could also be metrics from additional segments, such as age or gender. The benefit of using metrics such as age and gender is that they are data that can be quickly collected using single questions. Furthermore these questions are commonly requested details for a majority of applications at install or software registration and therefore users may be more likely to willingly divulge such personal information.

In addition to new metrics, the experimental application should be changed to one which does not use sound as the primary output mode, as in the web radio. This is because the time spent listening to the web radio may be considered passive interaction as opposed to active interaction with the application and as such confounds the data when analysing session data. An application whose interaction modes were primarily visually and keyboard/mouse based, or an independent logging mechanism as in Khan et al. (2008) would help to diminish this effect by reducing the ambiguity in active or passive interaction during a session. A web browser is suggested as a more appropriate software application in light of the findings here. One of the main factors in the decision to use a web radio was the fact that it offered a relatively limited functionality via any potential user interface, and as such sufficient diversity to allow any potential expression of personality trait in interactive behaviour without being overtly complex. A web browser offers similarly limited interactive options but sufficient opportunity for a range of interactive behaviours to be expressed, such as total clicks, session time, pages viewed, time per page yet without the auditory output issues of a web radio. Furthermore, the open source development of a number of popular web browsers means that logging functionality can be added via software modules known as extensions, providing an opportunity to investigate log file psychometrics without requiring significant resources.

However, despite these limitations there is evidence from this experiment to support the experimental hypothesis. Such a methodology for the deduction of personality from interactive behaviour expressed in log files offers an ability to provide a more personally relevant interaction using personality-based user interface skins for a potentially large anonymous user population, as stated in Chapter 2 in the segmentation section. This is because even with relatively low predictive accuracy to determine personality, it still provides a more personally relevant interaction. Furthermore, the predictive ability demonstrated by a test case shown in Chapter 4 illustrates prediction were for a large part correct and that even when the prediction was not correct there were a low number of major errors compared to minor errors, with the low consequence of minor errors still representing a more personally relevant, or personalised interaction compared to a single user interface approach.

A recommendation for future work in the particular area relating to H_1 is that researchers investigate this hypothesis in relation to their own specific applications, for example web or mobile application usage, because logically it follows that if you change the application you change the interaction. Therefore it also follows that one would need new application-specific metrics for the new application, although it is not possible to determine this until there is diverse application of this experimental methodology. If future work in this area uses this experimental methodology in other domains, a greater understanding of the nature of personality-relevant interaction metrics can be determined. There may be application-independent behavioural metrics that are indicators of personality, but as this is early research in this field it will require other applications of the methodology to better understand the nature of application-specific and independent personality- relevant behavioural metrics.

9.3.2 H_2: User personality can predict preference for user interface skin colour

Evidence to support the experimental hypothesis was found. The analysis demonstrated that according to this data, personality trait was a factor that influenced skin preference for colour. Future work to further understand the relationship between personality and preference for user interface colour could investigate the properties of colour, such as hue, saturation and lightness and their effect on preference for colour. Furthermore future work could focus on which particular colours are linked with which specific personality traits.

9.3.3 H_3: User personality can predict preference for user interface shape

No evidence to support the experimental hypothesis was found. Evidence to support a decline in skin preference as a function of the number of turns in the polygon was found although no links to personality trait could be found in this data. The irregular polygons used were not representative of a typical user interface shape (square, rectangular) and as such were possibly inappropriate for the transition between the paper-based polygons from the originator, Martindale et al. (1990) and as user on-screen interface skin shapes.

9.3.4 H_4: Users prefer user interface skins that exhibit a similar personality to their own

Evidence to support the experimental hypothesis was found for extraversion in supporting the Similarity Attraction Hypothesis. However, strong contradictory evidence for neuroticism was also found that suggested the inverse, a dissimilarity attraction possibly explained by aspiration or equilibrium seeking mechanisms. The third factor investigated, conscientiousness, provided internally contradicting results to one another that suggested that further research is required to understand whether similarity attraction applies to conscientiousness and its facets, and if so, under what circumstances. It is possible that there are attraction levels with different criteria for either similarity or dissimilarity attraction depending where on the personality trait scale the participant is, but this will need to be determined by further research. The experimental hypothesis has been revised for future work to include similarity attraction as a function of personality trait and position on the trait scale.

9.3.5 Overall Experimental Hypothesis: A more personally relevant interaction can be produced by using profiled user interface skins

This thesis has presented a means to provide a more personally relevant interaction using profiled user skins. Firstly a means to identify personality trait unobtrusively has been demonstrated through log file analysis. Secondly, three different aspects of user interface design elements have been investigated that might accommodate different personality traits: colour, shape and meaning. Links between personality, colour and meaning have been established, demonstrating that design for personality trait will likely produce a more preferable user interface skin when applied for the appropriate user, producing a more personally relevant interaction and therefore providing evidence to support the overall experimental hypothesis.

9.4 Reformulation of the Experimental Hypotheses for Future Work

This thesis has demonstrated a method for personalising interaction using user interface skins that have been designed for a particular personality profile. The hypotheses investigated by this thesis were stated in general terms, to reflect the

exploratory nature of this work. Future work could redefine these experimental hypotheses to test specific aspects of each of the hypotheses investigated here.

H_{1a} *Additional factors can be used to increase the variance in personality trait accounted for by interactive behaviour*

H_{1b} *Interactive behaviours can indicate personality trait independently of the application*

This hypothesis can be stated more specifically in terms of additional factors. The segments discussed in Chapter 2 may provide these additional factors to investigate, including Age and Gender. A second hypothesis investigates the nature of application-independent logging to determine Personality, to examine whether log files need to be understood within the context of the generating application.

H_{2a} *User personality can predict preference for skin hue*

H_{2b} *User personality can predict preference for skin saturation*

H_{2c} *User personality can predict preference for skin lightness*

Having established that Personality is a factor in skin colour preference, an interface design element that can be modified to appeal to a particular personality trait, elements of colour itself might be investigated such as Hue, Saturation and Lightness. The colour skins used in this research were controlled for saturation at 50%. Future hypotheses could be stated in terms of other colour components.

H_{3a} *User personality can predict preference for skin shape*

H_{3b} *User personality can predict preference for relative skin size*

The lack of support for the experimental hypothesis was likely due to the "outlandish" appearance to participants of the high number of sided polygons. As such future research should use user interface shapes that are more appropriate to be applied to user interface skins, such as rectangles that use workspace more efficiently and are more familiar to participants as user interfaces. The experimental hypothesis, H_3, is not restated because future work should retest this hypothesis with different shape

skins as described. In addition, future work could investigate the effect that personality has on preference for the relative user interface skin size on the desktop.

H_{4a} User preference for personality in a user interface varies as a function of personality trait and position on the personality trait scale

As detailed in Chapter 7, this hypothesis has been restated to include the concept of trait level within the hypothesis. The intention of this hypothesis is to investigate the conditions under which similarity or dissimilarity attraction occurs.

By developing the experimental hypotheses it is hoped that future work in this area can replicate and expand the findings presented here so that user interfaces can be designed for particular personality traits and therefore providing a more personalised interaction and away from average.

9.5 Contribution

9.5.1 Theoretical Contribution

This section expresses the contribution made by this work. This thesis has investigated four experimental hypotheses in determining whether a more personalised interaction can be produced using profiled user interface skins. As a result of these multiple investigations in search of a personalisation methodology, there are multiple contributions proposed by this work.

Perhaps the foremost contribution by this thesis is that personality as a construct is an HCI-relevant factor of use in the design of interaction. There is support for this from H_1, showing that interactive behaviour varies as a function of personality, H_2, showing that personality is a factor in skin preference and H_4, showing that personality can be expressed in skins, and that preference for skins varies as a function of the personality expressed in the skin and the personality of the user. Such support from multiple hypotheses indicates that, on the basis of this data, personality plays a role in interaction. The practical implication of this is that interaction designers can design for specific personality traits in order to provide a more personally relevant and personalised user experience.

In addition to the value of personality in interactive design, another contribution made by this work is the indirect linking of user behaviour to skin preference via a mediating factor (or subset) such as gender, age or personality. Typically linking user behaviour to skin preference directly might provide better predictive success but it does not provide as much insight on how to design skins for individual differences as much as via a mediating factor. By introducting this mediating factors, identification and targeting of user segment can be studied separately. For example once designers have established that they have an extrovert user segment that they want to target, they should be able to look at existing knowledge on how to design user interface skins for extroverts. It is possible in future developments within this field that specific design guidelines or design patterns will be established for designing for specific user segments. It will then be possible to use this knowledge over and over again across many different applications.

The next contribution made by this thesis is the exploration of a novel measure of personality from interactive behaviour, referred to by this work as log file psychometrics (LFP). As personality has been demonstrated to be an interaction design relevant construct, an ability to unobtrusively measure personality is preferable. LFP offers interaction designers this ability.

Another contribution made is the evidence to support Reeves and Nass' Computers as Social Actors (CASA) theory, also referred to by them in broader terms as "The Media Equation" that people interact with computers in similar ways as they do with other humans. The application of Byrne and Nelson's Similarity Attraction Hypothesis to HCI by Reeves and Nass has been tested and strong evidence found to suggest that the proposed Similarity Attraction Hypothesis does not hold true for all five factors, and that specifically for neuroticism that there appears to be a dissimilarity attraction. This further investigation and refinement of the Similarity Attraction Hypothesis in HCI represents a contribution by this work.

9.5.2 Practical Contribution

In addition to these contributions to theory there are two practical contributions suggested by this work. Firstly the jury-based methodology for establishing personality in user interface skins, investigated by H_4, provides a practical way for

designers to validate their personality-based designs to ensure that the same trait intended by the designer is the same trait perceived by users. A second practical contribution arising from this work is the use of an Internet-based platform to facilitate data communication between investigator and participants. This creates a feedback loop that may be of benefit to designers because it provides an understanding of the nature of a particular user group and then provides feedback in response to new user interface skin designs deployed to users, in the form of skin-specific interactive behaviours.

9.6 Limitations

A limitation of this study is the predominantly self reported data from the skin preference questionnaire used in the investigation of H_2, H_3 and H_4. It was the intention to use behavioural data as expressed preference (e.g. favourite skin by highest frequency of use) but insufficient behavioural data was generated and so skin preference data was used.

One of the main limitations of externalised research of this nature is that whilst the desired interactive behaviours can be captured and recorded, the surrounding context of use represents a number of uncontrolled variables that may or may not influence the interactive behaviours being observed. For example, with specific reference to this research, the web radio application provided a means for users to listen to Internet radio stations. However, as the context of use for each participant was uncontrolled the means for listening was unknown and could have been via speakers or headphones. This would likely have affected volume changing behaviour as a function of means for listening because of the greater impact that environmental sounds have on interrupting the user's listening experience when using speakers than they do when listening via headphones. In addition to volume changing behaviour there may have been environmental influences responsible for interrupting user listening, for example a telephone call or business meeting scheduled, that would not have been visible to the experiment. However, observing natural behaviour in the actual context of use may also be considered a strength of this type of research because it could be argued that unlike observation lounge studies, this externalised methodology captures natural, "real world" behaviours. Therefore, conclusions

drawn as a result of externalised research might be considered to have greater ecological validity than those drawn as a result of observation lounge studies.

One of the unexpected effects observed during the data collection was the contribution of knowingly wrongful data, possibly with malicious intent. There were 3 cases of users with originating IP addresses in Australia who completed the questionnaires in suspiciously quick times for reported 60 year old males with a preference for the pink skin of the colour skin pack (e.g. 60s completion time compared to mean of 400s). These were the only that cases of apparently valid data (i.e. complete) that were removed from the data set, although it raised the issue of participant integrity. The only checks that were performed in this research were the speed of completion of questionnaires, to ensure that unrealistic completion times could be discounted (which represented people who rapidly clicked on questionnaire answers without thinking to complete it quickly). Further research could establish a means for automatically checking for the integrity of contributed data.

Another limitation of this study is that there has been no control for cultural factors. Participants were obtained in part via the Internet from a project website which had visitors from all around the world. Similiarly the reported locations of participants in the general questionnaire indicated participation from all over Europe, Asia and Australasia. Cultural variance for colour, shape and meaning could affect external validity and further work needs to be done to determine if there is an effect, and whether such research should be controlled for geographically, or culturally.

9.7 Reflections

This section contains reflections on the contributions with the intention of providing the reader with additional insights learned by the experimenter during the course of this work. The section is comprised of two subsections: reflections on the theoretical and practical contributions of this thesis.

9.7.1 Reflections on Theoretical Contribution

This research has suggested log file psychometrics as a means to establish personality type from logged interaction data, however, once practically applied, log file recording itself becomes subject to privacy issues and laws. It is a well researched psychological phenomenon that human behaviour changes when knowingly observed, reffered to as the observer effect. For example the Hawthorne effect, where Mayo (1949) noted that productivity increased when workers were being watched. Mayo concluded that the behavioural change was due to a sense of feeling special. In these experiments it is possible that the opposite was happening – that an inverse Hawthorne effect was caused by a fear of log file recording. Such a fear would be based upon the participant's level of trust in the experimental application only logging the data it claimed to. In the current world context of prevalent Internet malware (virii, worms, rootkits, trojans etc.), user suspicion of privacy intrusions by applications creates a level of resistance to participation, and presents a greater challenge for research in this area. When canvassing for participation there was feedback to the experimenter expressing privacy concerns from number of potential participants. Furthermore, general participation numbers were lower than expected despite full transparency and of the application code, which seems like it was not part of the decision making process for the majority of potential participants. It was not learned during the experiments what type and level of information potential participants needed to feel sufficiently confident in the protection of their privacy to participate. However, it seems from this work that privacy issues are a concern for participants and therefore is an issue for this kind of externalised methodology.

9.7.2 Reflections on Practical Contribution

Perhaps the largest challenge to be faced during the research was generating large numbers of participants. Typically generating participants can be difficult so generating large numbers was expected to be challenging. However, a relatively large scale campaign to attract participants is not typical and so the aim was to attract approximately 1000 participants as an achievable figure. The reality was that 64 valid datasets were achieved. There are a number of factors that could account for this:

- Experimenter Bias: a dislike of the experimenter by participants. As an Internet radio station DJ promoting the research, it is possible that listeners did not like the voice or personality of the experimenter. One of the radio stations was gaming related and so any perception of the experimenter by other gamers would be within a competitive and combative gaming environment. The perception of this 'gaming personality' could well have affected people's decision as to whether they wanted to participate or not.

- Privacy Concerns: as noted in the previous section, any application that claims to log data seems to be handled with additional suspicion by users, creating additional barriers to participation (i.e. lack of trust).

- Commercial Competition: the experimental application was an Internet radio and as such needed to compete with other commercial Internet radio services (e.g. iTunes, Wunder Radio, Live365.com etc.) and media players (e.g. iTunes, WinAmp, Windows Media Player etc.).

- Extended Participation: the methodology involved participants using the experimental application for a period measured in weeks. This is asking participants to modify their regular Internet radio behaviours for an extended period of time, as compared to typically shorter observational studies. For example, if they did not like the radio player itself then it would have likely caused higher drop out rates because they would have been required to use it for the duration of the experiment (potentially 1.5 weeks). The same is likely true of the user interface skins that were transferred to participants every 3 days if conditions had been met. If the user did not like a particular skin pack (e.g. shape) then they would potentially have had to see those skins for a minimum of 3 days. One of the conditions for progression was a minimum number of radio player uses, and if the user didn't like the particular skinpack may not have used it, which in turn would have extended the time until the conditions had been met. This could have caused a tendency for participants with certain personality trait profiles to drop out (e.g. low conscientiousness).

The use of a jury to establish meaning in skins has been put forward as a contribution of this work. On reflection it was an effective means to apply a level of objectivity to

a subjectively perceived image. Practically however, there were methodological issues that could be improved upon. The main issue was the length of questionnaire designed and the time required to complete it. Ideally, each jury member would complete an IPIP-NEO personality inventory on each skin, for each of the four environmental questions (e.g. "I would use this skin when sitting at my desk") because it is correlates well with the well researched NEO-PIR (Costa and McCrae, 1985) However, the full IPIP-NEO instrument has 300 items in it and even the short version has 120 items and as such is difficult to administer repeatedly for several skins because of the extended time it takes to complete, which not only challenges the patience of most jury members but also makes recruiting them harder. As a result of this the TIPI inventory was used, and whilst it correlates well with IPIP-NEO for a ten item inventory, it may well be that it is not sophisticated enough for this type of application. Further research into a shorter item personality inventory than the IPIP-NEO inventories but longer than TIPI could help provide greater resolution than the TIPI can without requiring extended periods of time for jury members to complete, as in the case of the IPIP-NEO.

Despite the issues faced and the failure to achieve high participation, the externalised methodology developed in this thesis allows data to be collected from the field, and that is perhaps one of the core strengths of this research. However, externalisation presents a number of issues that have just been discussed. Future research should investigate externalisation as a methodology separately from hypotheses that require high numbers of participants so that externalisation itself as a methodology can be developed further.

9.8 Final Remarks

One of the challenges facing HCI is to address the issue of how to personalise interaction, in spite of the high diversity in individual differences. Reducing large user populations into smaller subsets defined by an individual difference, such as personality, provides designers the ability to provide skins specifically designed to appeal to particular personality traits. In so doing this moves interaction away from average and towards a more personally relevant interaction.

Furthermore, Internet-based research provides an ability to observe natural behaviours in the actual context of use. It provides accessibility and convenience for potential participants and therefore also has the potential for far higher numbers of participants than more traditional situated experimentation. If users are willing to give interaction data to researchers in the form of non-personally identifiable log files then researchers will be able to better understand the nature of interaction in general, and specifically how personality can be used to provide personalised interaction.

One of the higher level questions of relevance to this research area is the role of familiarity versus personalisation. Some users tend to prefer familiarity in the user interface, despite what might be considered to be inefficient or poorly designed user interfaces. Other users seek out ways to customise the interface, sometimes so extremely that they are in fact reducing their interactive efficacy in a trade off for increased aesthetic effect. The generational difference in interactive behaviours noted by Docampo Rama (2001) may also influence user tendency towards familiarity or personalisation, with older people preferring familiarity and younger people preferring personalisation.

One potential future direction for personalisation of the user interface is the portability of one's own personal UI across different applications, for example in your mobile phone, in-car navigation system, interactive digital television and home electronics (television, hi-fi, digital video recorder, satellite/cable, microwave, dishwasher etc). This provides a level of personalisation and familiarity and is device independent. Such portable user interfaces are similar to the UIMS architectures discussed in chapter 2, with one interface for multiple applications as opposed to the separable user interface approach that suggested multiple user interfaces for a single application. It might be that the UIMS architectures of the past provide the inspiration for future development in the personalisation of the user interface by combining the ideas of Separable User Interface theory with UIMS. For example, the UIMS architecture presents a single UI for all users across all platforms. By using the personalisation aspects of UI architectures from Separable User Interface theory a personalised user interface that can be applied to multiple applications can be achieved, providing the user consistency between applications and retaining the personally relevant aspects of the interface and subsequent interaction.

References

Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd Ed.). Milton-Keynes, England: Open University Press (McGraw-Hill).

Ali, M.R (1972). Pattern of EEG Recovery Under Photic Stimulation by Light of Different Colours, *Electroencephalography Clinical Neurophysiolology, 33,* p. 332-335.

Ambrose, G. A. and Harris, P. (2005). Colour, *Basic Design Series*, AVA Academic Publishing SA, Switzerland, p. 118-119.

Barrett, D. M. and Eaton, E. B. (1947). Preference for color or tint and some related test data. *Journal of Personality, 15,* 222-232.

Beckwith, L., Burnett, M., Wiedenbeck, S., Cook, C., Sorte, S. and Hastings, M. (2005). Effectiveness of End-User Debugging Software Features: Are There Gender Issues?, in *Proc. CHI 2005*, ACM Press (2005) 869-878

Benyon, D. R. (1993). Accommodating Individual Differences through an Adaptive User Interface. In *Adaptive user interfaces - Results and prospects*, Schneider-Hufschmidt, M., Kühme, T. and Malinowski, U. (eds.) Elsevier Science Publications, North-Holland, Amsterdam

Benyon, D. R. and Murray D. M. (1993). *Adapting systems to individual aptitudes. In proceedings of 1st international workshop on intelligent user interfaces*, Gray, W., Heffley, W. and Murray, D. M. (eds.) ACM Publications, New York.

Birnbaum, M.H. (2000). Psychological experiments on the Internet. San Diego, CA: Academic Press.

Blom, J. 2000. Personalization: a taxonomy. In *Chi '00 extended abstracts on human factors in computing systems* (The Hague, The Netherlands, April 01 - 06, 2000). New York: ACM Press, 313-314.

Blom, J.O., and Monk, A.F. (2003). Theory of personalization of appearance: Why users personalize their PCs and mobile phones. *Human-computer interaction, 18,* 193-228.

Brebner, J. (1983). A model of extraversion. *Australian Journal of Psychology, 35,* 349-359.

Brinkman, W.-P., & Fine, N. (2005). Towards Customized Emotional Design: An Explorative Study of User Personality and User Interface Skin Preferences. *Proceedings of EACE 2005*, 107-114.

Brinkman, W.-P., Haakma, R., & Bouwhuis, D.G. (2001). Usability evaluation of component-based user interfaces. In M. Hirose (Ed.), *Proceedings of IFIP INTERACT'01* (p.767-768). Amsterdam: IOS Press.

British Psychological Society (2006). *Code of Ethics and Conduct*, British Psychological Society, UK

Brown, A., & Logan, C. (2006). *The psychology of the Simpsons*. Dallas, TX: Benbella Books.

Buchanan, T. (1998). Internet research: Self-monitoring and judgements of attractiveness. Presented at the *28th annual convention of the Society for Computers in Psychology*, Dallas, TX.

Buchanan, T. and Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90,* 125-144.

Buchanan, T., Johnson, J. A., and Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the internet. *European J. of Psychological Assessment, 21,* 115-127.

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology, 1,* 659-663.

Cattell, R. B. (1950). *A systematic theoretical and factual study.* New York: McGraw-Hill.

Cattell, R. B. (1965). *The scientific analysis of personality.* Baltimore: Penguin Books.

Cerbus, G. and Nichols, R.C. (1963). Personality variables and response to color. *Psychological Bulletin*, 60, 566-575.

Chatzis, S., Doulamis, A., and Varvarigou, T. (eds.). (2007). A content-based image retrieval scheme allowing for robust automatic personalization. Proceedings from the *6th ACM international Conference on Image and Video Retrieval, Amsterdam, The Netherlands.* New York: ACM, 1-8.

Cooper, A. (1999). *The inmates are running the asylum.* New York: Macmillan Press.

Cooper, A. and Reiman, R. (2003). *About Face 2.0 – The Essentials of Interaction Design*, New York: Wiley.

Coles, G. H. (1989). SPR Presidential Address, 1988. *Psychophysiology, 26,* 251-269.

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory.* Odessa, FL: Psychology Assessment Resources.

Czaja, S. AND Lee, C. C. (2003). Designing computer systems for older adults. In J. A. Jacko and A. Sears (eds). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications.* Mahwah, NJ: Lawrence Erlbaum Associates, 413–427.

Czaja, S.J. and Sharit, J. (1998). Age differences in attitudes toward computers (1998) *Journal of Gerontology Psychological Sciences, 53B*, 5, 329 – 340.

Czerwinski, M., Tan, D. S. and Robertson, G. G. (2002). Women take a wider view. In *Proc. CHI 2002*. New York: ACM Press, 195-202.

De Sousa, L. G. and Leite, J. C. (2003). XICL: a language for the user's interfaces development and its components. Proceedings from *CLIHC '03: The Latin American conference on Human-computer interaction. Rio de Janeiro, Brazil*. 191-200.

Docampo Rama, M., (2001). Technology generations handling complex user interfaces. *Doctoral dissertation, Technische Universiteit Eindhoven*, The Netherlands.

Doucet, C. and Stelmack, R.M. (1997). Movement time differentiates extraverts from introverts. *Personality and Individual Differences, 23,* 775-786.

Duffy, E. (1962). *Activation and behaviour.* New York: Wiley.

Dufresne, A., & Turcotte, S. (1997). Cognitive style and its implications for navigation strategies. In B. Boulay & R. Mizoguchi (ed.). *Artificial intelligence in education knowledge and media learning system.* Kobe, Japan: Amsterdam IOS Press, 287-293.

Edmonds, E.A. (ed.). (1992). *The Separable User Interface.* San Diego, CA: Academic Press, Inc.

Encelle, B. and Baptiste-Jessel, N. (2007). Personalization of user interfaces for browsing XML content using transformations built on end-user requirements. In proceedings from the 2007 international *Cross-Disciplinary Conference on Web Accessibility (W4a) Banff, Canada.* New York: ACM 58-64.

Evers, M. (1999). *A case study on adaptability problems of the separation of user interface and application semantics*. CTIT Technical Report TR99-14, Centre for Telematics and Information Technology.

Eysenck, H. (1961). The effects of psychotherapy: An evaluation. In H. J. Eysenck (ed.). *Handbook of Abnormal Psychology: An experimental approach*. New York: Basic Books, 697-725.

Eysenck, H. (1991). Personality, Stress and Disease: An interactionist perspective. *Psychological Inquiry, 2,* 221-232.

Eysenck, H. J. (1941) Type factors in aesthetic judgements. *British Journal of Psychology, 31,* 262-270.

Eysenck, H. J. (1967) *The biological basis of personality*. Springfield, IL: Thomas Publishers.

Eysenck, H.J. & Eysenck, S.B.G. (1976). *Psychoticism as a dimension of personality*. London: Hodder and Stoughton.

Eysenck, H.J. (1947). *Dimensions of personality.* London, U.K.: Routledge & Kegan Paul.

Ferman, A. M., Errico, J. H., Beek, P. v., and Sezan, M. I. (2002). Content-based filtering and personalization using structured metadata. In Proceedings of the *2nd ACM/IEEE-CS Joint Conference on Digital Libraries, Portland, Oregon, USA*. New York: ACM. 393-393.

Fine, N., & Brinkman, W.-P. (2006). Towards extracting personality trait data from interaction behaviour. In W.-P. Brinkman (ed.), *Proceedings of the 2006 Workshop on Computer Assisted Recording, Pre-Processing, and Analysis of User Interaction Data,* (p. 76-81). Morrisville, NC: Lulu.

Foley, J (1991). Future Directions in User-Computer Interface Software, *ACM SIGOIS Bulletin, Conference proceedings on Organizational computing systems,12,* 2-3.

Fox, S., Karnawat, K., Mydland, M., Dumais, S., and White, T. (2005). Evaluating implicit measures to improve web search. *ACM Trans. Inf. Syst. 23* (2), 147-168.

Furnham, A. (1997). *The psychology of behaviour at work.* Hove: Psychology Press.

Geen RG. (1984). Preferred stimulation levels in introverts and extroverts: Effects on arousal and performance. *J Personal Soc Psychol, 46*,1303-1312.

Gena, C. (2001). Designing TV Viewer Stereotypes for an Electronic Program Guide. In *Proceedings of the 8th international Conference on User Modeling 2001*. M. Bauer, P. J. Gmytrasiewicz, and J. Vassileva (Eds.). *Lecture Notes In Computer Science, 2109*. London: Springer-Verlag, 274-276.

Gibson, J. J. (1979). *The ecological approach to visual perception.* New Jersey: Lawrence Erlbaum Associates.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five Factor structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe, 7*, 7-28. Tilburg, The Netherlands: Tilburg University Press.

Goldstein, M., Alsiö, G., and Werdenhoff, J. (2002). The media equation does not always apply: people are not polite towards small computers. *Personal Ubiquitous Computing 6*, 2, 87-96.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the big five personality domains. *Journal of Research in Personality, 37*, 504-528.

Granger, G.W. (1955). The prediction of preference for color combinations. *Journal of General Psychology, 52*, 213-222.

Green M. (1985). Report on dialogue specification tools. In: *User Interface Management Systems*. Pfaff, G.E. (Ed.). Berlin: Springer Verlag.

Grillo, P. (1975). *Form, Function and Design*, New York: Dover Publications.

Haralambos, M., Holborn, M. and Heald, R. (2006). *Sociology: Themes and Perspectives*, London: Harper Collins Publishers.

Hassenzahl, M. (2004). The interplay of beauty, goodness and usability in interactive products. *Human Computer Interaction, 19,* 319-349.

Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting with Computers 12,* Elsevier Science B.V, 507-528

Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system), *Psychological Review, 62,* 243-254.

Helve, J. and Krause, U. (1972). The influence of age on performance in the Panel-D15 colour vision test, *Acta Opthalmologica 50,* 896-901.

Hilbert, D. M. and Redmiles, D. F. 1998. An approach to large-scale collection of application usage data over the Internet. In *Proceedings of the 20th international Conference on Software Engineering, Kyoto, Japan.* IEEE Computer Society, Washington, DC, 136-145.

Hilbert, D. M. and Redmiles, D. F. (2000). Extracting usability information from user interface events. *ACM Comput. Surv. 32,* 4, 384-421.

Hoashi, K., Matsumoto, K., and Inoue, N. (2003). Personalization of user profiles for content-based music retrieval based on relevance feedback. In *Proceedings of the Eleventh ACM international Conference on Multimedia, Berkeley, CA, USA.* New York: ACM, 110-119.

Holmes, C. B. and Buchanan, J. A. (1984). Color preference as a function of the object described. *Bulletin of the Psychonomic Society, 22* (5), 423-425.

Hurley, W.D. and Sibert, J.L. (1992). Modelling User Interface-Application Interactions, in: E. Edmonds (ed.), *The separable user interface*, Academic Press, pp. 151-165.

Isbister, K. (2005). *Better game characteristics by design: a psychological approach,* Morgan Kaufman Publishers Inc.

Johnson, D.L., Weibe, J.S., Gold, S.M., Andreasen, N.C, Hichwa, R.D., Watkins, G.L. and Boles Ponto, L.L. (1999). Cerebral blood flow and personality: a positron emission tomography study. *American Journal of Psychiatry, 156*, 252-7.

Kelly, D. and Teevan, J. (2003). Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum 37*, 2, 18-28.

Kline, D.W. and Scialfa, C.T. (1996). Sensory and perceptual functioning: basic research and human factors implications, in A.D. Fisk, W.A. Rogers (eds), *Handbook of Human Factors and the Older Adult*, San Diego, CA: Academic Press.

Kreitler, H. and Kreitler, S. (1972). *Psychology of the Arts.* Durham: Duke University Press.

Kroemer, K. H. E. and Grandjean, E. (1997). *Fitting the task to the human* (5th ed.). London, UK: Taylor & Francis.

Kunce, J.T. and Tamkin, A.S. (1981). Rorschach movement and color responses and MMPI social extraversion and thinking introversion personality types. *J. of Personality Assessment, 45*, 5-10.

Kurosu, M. and Kashimura, K., (1995). Apparent usability vs. inherent usability, *CHI*, 292-293.

Lavie, T. and Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites, *International Journal of Human-Computer Studies, 60*(3):269-298.

Liffick, B. and Yohe, L. (2001). Using Surveillance Software as an HCI Tool, In *Proceedings of Information Systems Education Conference (ISECON).*

Limbourg, Q., Vanderdonckt, J., Michotte, B., Bouillon, L., Lopez-Jaquero, V. (2004). UsiXML: a Language Supporting Multi-Path Development of User Interfaces. In: *Proc. of EHCI-DSVIS'2004,* Dordrecht: Kluwer Academics, 207-228.

Mandler, G. (1992). Memory arousal and mood: A theoretical integration. In S. Christianson (Ed.). *The handbook of emotion and memory: Research and theory.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Matthews, G., Deary, I.J., & Whiteman, M.C. (2003). *Personality traits* (2nd edition). Cambridge: Cambridge University Press.

Mayo, E. (1933). *The human problems of an industrial civilization,* New York: MacMillan.

Mayo, E. (1949). *Hawthorne and the Western Electric Company, The Social Problems of an Industrial Civilisation,* Routledge.

Mayora-Ibarra, O., de la Paz-Arroyo, O., Cambranes-Martínez, E., and Fuentes-Penna, A. (2003). A visual programming environment for device independent generation of user interfaces. In *Proceedings of the Latin American Conference on Human-Computer interaction, Rio de Janeiro, Brazil. 46*, New York: ACM Press, 61-68.

Mischel, W. (1993). *Introduction to Personality*. Fifth Edition. Fort Worth, USA: Harcourt Brace Jovanovich.

Moon, Y. and Nass, C. I. (1996). Adaptive agents and personality change: complementarity versus similarity as forms of adaptation. In *Conference companion on human factors in computing systems: Common ground*, Vancouver, British Columbia, Canada. M. J. Tauber, Ed. New York: ACM Press, 287-288.

Munsinger, H.L. and Kessen, W. (1964). Uncertainty, structure and preference, *Psychological Monographs, 78,* 586.

Myers, B. (1998). A brief history of human computer interaction technology. *ACM interactions, 5* (2), 44-54.

Myers, B.. Hudson, S.E., and Pausch, R. (2000). Past, present, and future of user interface software tools, *ACM Transactions on Computer-Human Interaction, 7* (1), 3-28.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating interdependence, Boston, Massachusetts*. Adelson, S. Dumais, and J. Olson (Eds.). New York: ACM Press, 72-78.

Nass, N., Moon, Y., Fogg, B., Reeves, B. and Dryer, C. (1995). Can Computer Personalities Be Human Personalities? In *Proceedings of CHI*.

Nielsen, J (1993) *Usability engineering*, Boston: Academic Press.

Norman, D. A. (2002). Emotion and design: Attractive things work better. *Interactions Magazine, ix* (4), 36-42.

Norman, D.A. (1988): *The design of everyday things*. New York, Doubleday.

O'Hare, D. (1981). *Psychology and the Arts*. Brighton: Harvester Press.

O'Gorman, J. G. (1977). Individual differences in habituation of human physiological responses: a review of theory, method, and findings in the study of personality correlates in non-clinical populations. *Biological Psychology, 5*, 257-318.

Omojokun, O. and C. Isbell (2003). User modelling for personalized universal appliance application interaction. In *Proceedings of Tapia*.

O'Neill, P., Roast, C., and Hawley, M. (2000). Evaluation of scanning user interfaces using real-time-data usage logs. In *Proceedings of the Fourth international ACM Conference on Assistive Technologies, Arlington, Virginia*. New York: ACM Press, 137-141.

Pasveer, K. A. and Ellard, J.H. (1998). The making of a personality inventory: Help from the WWW. *Behaviour Research Methods, Instruments and Computers, 30,* 309-313.

Phanouriou, C. (2000). *UIML: A Device-Independent User Interface Markup Language,* PhD thesis, Department of Computer Science, Virginia Tech, USA.

Pruitt, J. and Grudin, J. (2003). Personas: practice and theory. In *Proceedings of the 2003 conference on designing for user experiences (San Francisco, California)*. New York: ACM Press, 1-15.

Puerta, A. and Eisenstein, J. (2002). XIML: a common representation for interaction data. In *Proceedings of the 7th international Conference on intelligent User*

interfaces San Francisco, California, USA. New York: ACM Press, NY, 214-215.

Reeves B. and Nass C. (1996). *The Media Equation. How people treat computers, television and new media like real people and places.* Stanford, CA: Cambridge University Press, CLSI Publications.

Reynolds, C. (1997). A critical examination of separable user interface management systems: constructs for individualization. *SIGCHI Bull. 29*, 3 (July), 41-45.

Rich E. (1989): Stereotypes and User Modeling. In A. Kobsa and W. Wahlster, (Eds.), *User Models in Dialog Systems*, pp. 31-51. Berlin: Springer Verlag.

Rosenberg, J., Hill, R., Miller, J., Schulert, A., and Shewmake, D. (1988). UIMSs: threat or menace? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Washington, D.C., USA*. J. J. O'Hare (Ed.) New York: ACM Press, 197-200.

Saari, T. and Turpeinen, M. (2004). Towards psychological customization of information for individuals and social groups. In *Designing Personalized User Experiences in Ecommerce*, J. Karat, J. Vanderdonekt, C. Karat, and J. O. Blom, (Eds.) Human-Computer Interaction Series, vol. 5. Norwell, MA: Kluwer Academic Publishers, 19-37.

Saati, B., Salem, M. & Brinkman, W.-P. (2005). Towards customized user interface skins: investigating user personality and skin colour, *Proceedings of HCI 2005*, vol. 2, 89-93.

Sheskin, D. J. (1997). *Parametric and Non Parametric Statistical Procedures*. CRC Press.

Siochi, A. C. and Ehrich, R. W. (1991). Computer analysis of user interfaces based on repetition in transcripts of user sessions. *ACM Trans. Inf. Syst. 9*, 4 (Oct), 309-335.

Siochi, A. C. and Hix, D. (1991). A study of computer-supported user interface evaluation using maximal repeating pattern analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology, New Orleans, Louisiana, United States*. S. P. Robertson, G. M. Olson, and J. S. Olson (Eds.) New York: ACM Press, 301-305.

Smith, M. A. and Leigh, B. (1997) Virtual subjects: Using the Internet as an alternative source of subjects and research environments. *Behaviour Research Methods, Instruments and Computers, 29,* 496-505.

Stimpson, D. V. & Stimpson, M. F. (1979). Relation of personality characteristics and color preferences. *Perceptual & Motor Skills, 49*, 60-62.

Stone, N. J. (2003). Environmental view and color for a simulated telemarketing task. *Journal of Environmental Psychology, 23*, 63-78.

Tractinsky, N. and Zmiri, D. (2005) Exploring Attributes of Skins as Potential Antecedents of Emotion in HCI, in Fishwick, P. (Ed.) *Aesthetic Computing*, MIT Press.

Tractinsky, N., Shoval-Katz A. and Ikar, D. (2000). What is Beautiful is Usable Interacting with Computers, 13(2): 127-145.

Valdez, P. and Mehrabian, A. (1994) Effects of colour on emotions. *Journal of Experimental Psychology, 123*, 4, December.p394-409.

van der Veer, G. C. (1989). Individual Differences and the user interface. *Ergonomics, 32,* 1431-1449.

van der Veer, G. C. (1990) *Human Computer Interaction: Learning, Individual Differences and Design Recommendations.* Alblasserdam: Offsetdrukkerij, Haveka, B.V.

Vicente, K.J. and Williges, R.C. (1988). Accommodating individual differences in searching hierarchical file systems. Inter*national Journal of Man-Machine Studies, 29,* 647-668.

Watson, D., & Walker, L. M. (1996). The long-term temporal stability and predictive validity of trait measures of affect. *Journal of Personality and Social Psychology, 70,* 567-577.

Wiecha, C., Bennett, W., Boies, s., and Gould, J. (1989). Generating highly interactive user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Wings For the Mind*, K. Bice and C. Lewis, (Eds.) New York: ACM Press, 277-282.

Witkin, H. A., Dijk, R. B., Faterson, H. F., Goodenough, D. R., & Kapp, S. A. (1962). *Psychological Differentiation: Studies of development*. New York: Wiley.

Wolfson, S. and Case, G. (2000). The effects of sounds and colour on responses to a computer game. *Interacting with Computers. 13,* 2, (December), 183-192.

Yerkes, R.M. ad Dodson, J. J. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482.

Zuckerman, M. (1979). *Sensation seeking: beyond the optimal level of arousal*. London: Wiley.

Internet References

Berkeley Open Infrastructure for Network Computing
http://boinc.berkeley.edu/

SETI@HOME
http://setiathome.ssl.berkeley.edu/

FOLDING@HOME
http://folding.stanford.edu/

LHC@home
http://lhcathome.cern.ch/

Distributed.net
http://distributed.net/

Appendices

Appendix A: Full MANCOVA output for Hypothesis 2, Chapter 5

MANCOVA (all subjects, n=64)

TRAIT	COLOUR				COLOUR AND TRAIT				TRAIT			
	F	df	error df	p	F	df	error df	p	F	df	error df	p
Extraversion	5.256	7	55	0.001	0.2	7	55	0.984	2.59	1	61	0.113
Friendliness	5.556	7	55	0.001	0.528	7	55	0.81	0.284	1	61	0.596
Gregariousness	5.617	7	55	0.001	0.241	7	55	0.973	1.668	1	61	0.201
Assertiveness	6.808	7	55	0.001	0.758	7	55	0.624	0.744	1	61	0.392
Activity Level	5.192	7	55	0.001	0.424	7	55	0.883	0.402	1	61	0.528
Excitement Seeking	14.846	7	55	0.001	3.933	7	55	0.002 *	5.228	1	61	0.026
Cheerfulness	5.748	7	55	0.001	0.201	7	55	0.984	2.685	1	61	0.106
Agreeableness	1.584	7	55	0.16	2.139	7	55	0.054	4.207	1	61	0.045
Trust	3.446	7	55	0.004	0.502	7	55	0.829	0.084	1	61	0.773
Morality	2.697	7	55	0.018	3.423	7	55	0.004 *	7.271	1	61	0.009
Altruism	3.674	7	55	0.003	0.664	7	55	0.701	1.588	1	61	0.212
Cooperation	2.583	7	55	0.022	5.139	7	55	0.001 *	6.449	1	61	0.014
Modesty	3.301	7	55	0.005	1.647	7	55	0.142	0.591	1	61	0.445
Sympathy	6.309	7	55	0.001	0.534	7	55	0.805	0.457	1	61	0.502
Conscientiousness	3.897	7	55	0.002	1.339	7	55	0.25	3.921	1	61	0.052
Self Efficacy	6.685	7	55	0.001	1.622	7	55	0.149	0.791	1	61	0.377
Orderliness	5.812	7	55	0.001	0.72	7	55	0.655	0.42	1	61	0.519
Dutifulness	3.269	7	55	0.006	1.692	7	55	0.13	3.161	1	61	0.08
Achievement Striving	5.193	7	55	0.001	1.5	7	55	0.186	1.721	1	61	0.194
Self Discipline	7.309	7	55	0.001	1.051	7	55	0.407	0.184	1	61	0.67
Cautiousness	3.428	7	55	0.004	2.484	7	55	0.027 *	2.551	1	61	0.115
Neuroticism	9.367	7	55	0.001	1.269	7	55	0.283	0.597	1	61	0.443
Anxiety	10.048	7	55	0.001	0.925	7	55	0.494	1.01	1	61	0.319
Anger	13.322	7	55	0.001	2.394	7	55	0.033 *	0.805	1	61	0.373
Depression	7.411	7	55	0.001	0.272	7	55	0.962	0.148	1	61	0.702
Self Consciousness	5.145	7	55	0.001	0.22	7	55	0.979	0.171	1	61	0.68
Immoderation	10.456	7	55	0.001	1.783	7	55	0.109	1.711	1	61	0.196
Vulnerability	7.328	7	55	0.001	1.16	7	55	0.341	0.118	1	61	0.732
Openness to new experience	4.68	7	55	0.001	1.517	7	55	0.181	0.077	1	61	0.782
Imagination	9.452	7	55	0.001	1.205	7	55	0.315	0.219	1	61	0.641
Artistic Interests	5.413	7	55	0.001	0.949	7	55	0.477	0.184	1	61	0.669
Emotionality	5.791	7	55	0.001	0.684	7	55	0.685	0.964	1	61	0.33
Adventurousness	3.812	7	55	0.002	1.346	7	55	0.247	0.211	1	61	0.648
Intellect	3.692	7	55	0.002	0.734	7	55	0.644	1.681	1	61	0.2
Liberalism	2.469	7	55	0.028	3.339	7	55	0.005 *	0.048	1	61	0.827

* indicates significant at <0.05%, ** indicates significant at <0.01%

H_2: Full list of MANCOVA results for the colour skin pack, all participants

MANCOVA (male subjects, n=55)

TRAIT	COLOUR				COLOUR AND TRAIT					TRAIT			
	F	df	error df	p	F	df	error df	p		F	df	error df	p
Extraversion	9.275	7	46	0.001	0.409	7	46	0.892		1.435	1	52	0.236
Friendliness	9.11	7	46	0.001	0.571	7	46	0.776		0.001	1	52	0.996
Gregariousness	9.809	7	46	0.001	0.424	7	46	0.882		0.945	1	52	0.336
Assertiveness	10.553	7	46	0.001	0.999	7	46	0.444		1.642	1	52	0.206
Activity Level	7.793	7	46	0.001	0.471	7	46	0.851		0.014	1	52	0.907
Excitement Seeking	18.181	7	46	0.001	2.942	7	46	0.012	*	11.483	1	52	0.001
Cheerfulness	8.498	7	46	0.001	0.289	7	46	0.955		2.774	1	52	0.102
Agreeableness	2.676	7	46	0.021	1.816	7	46	0.107		8.538	1	52	0.005
Trust	5.103	7	46	0.001	0.554	7	46	0.789		0.297	1	52	0.588
Morality	4.528	7	46	0.001	2.848	7	46	0.015	*	10.852	1	52	0.002
Altruism	7.002	7	46	0.001	0.519	7	46	0.815		3.459	1	52	0.069
Cooperation	2.526	7	46	0.028	4.104	7	46	0.001	**	14.243	1	52	0.003
Modesty	4.829	7	46	0.001	1.087	7	46	0.387		0.031	1	52	0.862
Sympathy	8.952	7	46	0.001	0.555	7	46	0.788		0.384	1	52	0.538
Conscientiousness	5.516	7	46	0.001	1.172	7	46	0.337		11.316	1	52	0.009
Self Efficacy	8.863	7	46	0.001	1.328	7	46	0.259		3.082	1	52	0.085
Orderliness	7.634	7	46	0.001	0.756	7	46	0.627		0.698	1	52	0.407
Dutifulness	4.988	7	46	0.001	1.811	7	46	0.108		4.84	1	52	0.032
Achievement Striving	7.443	7	46	0.001	1.499	7	46	0.192		3.407	1	52	0.071
Self Discipline	9.938	7	46	0.001	1.335	7	46	0.256		1.226	1	52	0.273
Cautiousness	4.402	7	46	0.001	1.938	7	46	0.085		4.446	1	52	0.04
Neuroticism	11.252	7	46	0.001	0.908	7	46	0.509		0.731	1	52	0.397
Anxiety	12.735	7	46	0.001	1.11	7	46	0.373		0.593	1	52	0.445
Anger	16.717	7	46	0.001	2.043	7	46	0.07		0.888	1	52	0.35
Depression	9.266	7	46	0.001	0.409	7	46	0.892		0.068	1	52	0.796
Self Consciousness	5.571	7	46	0.001	0.376	7	46	0.912		0.012	1	52	0.912
Immoderation	15.993	7	46	0.001	2.101	7	46	0.062		3.472	1	52	0.068
Vulnerability	8.449	7	46	0.001	0.724	7	46	0.652		0.212	1	52	0.647
Openness to new experience	8.193	7	46	0.001	1.35	7	46	0.249		0.339	1	52	0.563
Imagination	12.892	7	46	0.001	1.337	7	46	0.255		1.138	1	52	0.291
Artistic Interests	8.501	7	46	0.001	0.908	7	46	0.509		0.013	1	52	0.908
Emotionality	9.337	7	46	0.001	0.508	7	46	0.824		2.215	1	52	0.143
Adventurousness	7.435	7	46	0.001	1.238	7	46	0.302		0.001	1	52	0.974
Intellect	6.147	7	46	0.001	0.946	7	46	0.481		2.342	1	52	0.132
Liberalism	3.874	7	46	0.002	2.705	7	46	0.02	*	0.023	1	52	0.881

* indicates significant at <0.05%, ** indicates significant at <0.01%

H_2: Full list of MANCOVA results for the colour skin pack, male participants

MANCOVA (female subjects, n=9)

TRAIT	COLOUR				COLOUR AND TRAIT				TRAIT			
	F	df	error df	p	F	df	error df	p	F	df	error df	p
Extraversion	0.886	7	1	0.677	1.029	7	1	0.643	1.661	1	7	0.238
Friendliness	0.448	7	1	0.821	1.424	7	1	0.57	2.995	1	7	0.127
Gregariousness	1.06	7	1	0.636	0.502	7	1	0.799	0.501	1	7	0.502
Assertiveness	0.459	7	1	0.817	0.976	7	1	0.655	0.645	1	7	0.448
Activity Level	4.485	7	1	0.349	6.163	7	1	0.301	5.721	1	7	0.048
Excitement Seeking	5.37	7	1	0.321	5.056	7	1	0.33	1.119	1	7	0.325
Cheerfulness	2.66	7	1	0.441	7.313	7	1	0.278	0.01	1	7	0.923
Agreeableness	0.564	7	1	0.775	0.241	7	1	0.919	1.215	1	7	0.307
Trust	337.49	7	1	0.042 *	447.49	7	1	0.036 *	2.942	1	7	0.13
Morality	0.331	7	1	0.874	0.43	7	1	0.829	0.679	1	7	0.437
Altruism	0.539	7	1	0.785	0.155	7	1	0.961	0.765	1	7	0.411
Cooperation	5.336	7	1	0.322	1.299	7	1	0.591	0.001	1	7	0.977
Modesty	1.563	7	1	0.55	0.832	7	1	0.691	3.256	1	7	0.114
Sympathy	0.129	7	1	0.973	0.253	7	1	0.913	0.009	1	7	0.925
Conscientiousness	1.908	7	1	0.507	1.35	7	1	0.582	2.276	1	7	0.175
Self Efficacy	3.651	7	1	0.383	33.766	7	1	0.132	2.705	1	7	0.144
Orderliness	9.492	7	1	0.245	2.665	7	1	0.44	0.022	1	7	0.866
Dutifulness	0.825	7	1	0.693	0.808	7	1	0.697	0.438	1	7	0.529
Achievement Striving	0.733	7	1	0.719	1.028	7	1	0.643	0.517	1	7	0.495
Self Discipline	211.521	7	1	0.053	409.038	7	1	0.038 *	4.676	1	7	0.067
Cautiousness	3.945	7	1	0.37	1.701	7	1	0.532	0.346	1	7	0.575
Neuroticism	0.777	7	1	0.706	2.574	7	1	0.447	0.135	1	7	0.725
Anxiety	0.283	7	1	0.898	0.568	7	1	0.774	1.784	1	7	0.223
Anger	0.439	7	1	0.825	0.621	7	1	0.755	0.071	1	7	0.797
Depression	5.269	7	1	0.324	4.25	7	1	0.358	0.639	1	7	0.45
Self Consciousness	3.801	7	1	0.376	3.254	7	1	0.403	1.24	1	7	0.302
Immoderation	0.271	7	1	0.904	1.446	7	1	0.567	1.335	1	7	0.286
Vulnerability	1.577	7	1	0.548	2.491	7	1	0.454	0.03	1	7	0.867
Openness to new experience	2.361	7	1	0.464	1.229	7	1	0.603	0.155	1	7	0.705
Imagination	3.304	7	1	0.401	5.451	7	1	0.319	3.315	1	7	0.111
Artistic Interests	0.613	7	1	0.758	0.293	7	1	0.893	0.279	1	7	0.614
Emotionality	0.641	7	1	0.748	3.484	7	1	0.391	0.975	1	7	0.356
Adventurousness	5.812	7	1	0.309	4.402	7	1	0.352	1.718	1	7	0.231
Intellect	6.088	7	1	0.303	1.214	7	1	0.606	0.002	1	7	0.965
Liberalism	4.173	7	1	0.361	1.24	7	1	0.601	1.414	1	7	0.273

* indicates significant at <0.05%, ** indicates significant at <0.01%

H_2: Full list of MANCOVA results for the colour skin pack, female participants

Appendix B: Summary table of all correlations for H_2, Chapter 5 showing acceptances of a priori and post hoc alphas

All Participants	Skin Colour	Self Reported	p (two tailed)
Factor		n=63	
Excitement Seeking	Grey	-0.425	**0.001**
	Yellow	-0.364	**0.003**
Agreeableness	Grey	0.332	*0.008*
	Purple	0.329	*0.008*
Morality	Grey	0.388	**0.002**
	Blue	0.298	*0.018*
	Purple	0.42	**0.001**
Cooperation	Grey	0.468	**0.001**
	Yellow	0.318	*0.011*
	Pink	-0.327	*0.009*
Modesty		-0.266	*0.035*
Conscientiousness	Grey	0.298	*0.018*
	Yellow	0.258	*0.041*
Dutifulness	Grey	0.322	*0.010*
Cautiousness	Grey	0.445	**0.001**
Neuroticism	Grey	-0.273	*0.031*
Anger	Grey	-0.379	**0.002**
Immoderation	Grey	-0.329	*0.009*
Openness to New Experience	Orange	0.267	*0.034*
Liberalism	Grey	0.327	*0.009*
	Red	-0.297	*0.018*

Male Participants	Skin Colour	Self Reported	p (two tailed)
Factor		n=54	
Assertiveness	Yellow	0.287	*0.036*
	Pink	0.332	*0.014*
Excitement Seeking	Grey	-0.398	**0.003**
	Red	-0.352	*0.009*
	Yellow	-0.436	**0.001**
Agreeableness	Grey	0.293	*0.031*
	Blue	0.268	*0.050*
	Orange	0.301	*0.027*
	Purple	0.383	**0.004**
Morality	Grey	0.391	**0.003**
	Blue	0.341	*0.012*
	Purple	0.456	**0.001**
Cooperation	Grey	0.453	**0.001**
	Red	0.273	*0.046*
	Yellow	0.298	*0.028*
	Blue	0.278	*0.041*
Conscientiousness	Red	0.341	*0.012*
	Yellow	0.334	*0.014*
Self Efficacy	Red	0.323	*0.017*
	Yellow	0.364	*0.007*
Dutifulness	Grey	0.319	*0.019*
	Blue	0.280	*0.040*
Achievement Striving	Red	0.335	*0.013*
	Yellow	0.280	*0.040*
Self Discipline	Red	0.328	*0.015*
Cautiousness	Grey	0.417	**0.002**
Anger	Grey	-0.397	**0.003**
Immoderation	Grey	-0.336	*0.013*
Openness to New Experience	Orange	0.293	*0.031*
Emotionality	Red	0.288	*0.035*
Liberalism	Grey	0.33	*0.015*

Female Participants	Skin Colour	Self Reported	p (two tailed)
Factor		n=9	
Extraversion	Green	-0.670	*0.049*
Friendliness	Grey	0.676	*0.046*
	Pink	-0.696	*0.037*
Activity Level	Red	-0.681	*0.043*
	Green	-0.749	*0.020*
Excitement Seeking	Blue	0.776	*0.014*
Agreeableness	Grey	0.773	*0.015*
Altruism	Grey	0.674	*0.047*
Modesty	Grey	0.811	*0.008*
	Orange	-0.759	*0.018*
Conscientiousness	Red	-0.676	*0.046*
	Green	-0.744	*0.021*
Self Efficacy	Green	-0.723	*0.028*
Self Discipline	Red	-0.677	*0.045*
	Green	-0.857	**0.003**
Imagination	Grey	-0.743	*0.022*
Trust	Red	-0.848	**0.004**

bold indicates acceptance under post hoc alpha = 0.0064

italics indicates acceptance under a priori alpha = 0.05