Evaluation of the Influence of Personality Types on Performance of Shared Tasks in a Collaborative Environment

Sinéad Mc Givney BSc. Hons.

A dissertation submitted in partial fulfilment of the requirements for the award of
Doctor of Philosophy

to the

Center For Digital Video Processing
School of Computing
Dublin City University

Supervisor: Prof. Alan F. Smeaton

September, 2008.

This thesis is based on the candidate's own work, and has not previously been submitted for a degree at any academic institution.

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others, save, and to the extent that, such work has been cited and acknowledged within the text of my work.

Signed:

ID No:

Date:

Abstract

Computer Supported Cooperative Work (CSCW) is an area of computing that has been receiving much attention in recent years. Developments in groupware technology, such as MERL's Diamondtouch and Microsoft's Surface, have presented us with new, challenging and exciting ways to carry out group tasks. However, these groupware technologies present us with a novel area of research in the field of computing – that being multi-user Human-Computer Interaction (HCI). With multi-user HCI, we no longer have to cater for one person working on their own PC. We must now consider multiple users and their preferences as a group in order to design groupware applications that best suit the needs of that group.

In this thesis, we aim to identify how groups of two people (dyads), given their various personality types and preferences, work together on groupware technologies. We propose interface variants to both competitive and collaborative systems in an attempt to identify what aspects of an interface or task best suit the needs of the different dyads, maximising their performance and producing high levels of user satisfaction. In order to determine this, we introduce a series of user experiments that we carried out with 18 dyads and analyse their performance, behaviour and responses to each of 5 systems and their respective variants. Our research and user experiments were facilitated by the DiamondTouch – a collaborative, multi-user tabletop device.

Acknowledgements

There are many people to whom I owe my sincere gratitude for supporting and helping me to achieve this PhD. First and foremost, I would like to thank my supervisor, Prof. Alan Smeaton, for giving me the opportunity to undertake this work and for his advice and guidance over the past few years. I would especially like to thank Dr. Hyowon Lee, who has always been there to offer advice, put forward ideas and answer any questions that I had. His help, guidance and friendship has been highly significant to my receipt of this doctorate.

Many thanks also go to all of the researchers at MERL, who provided access to the DiamondTouch and made my research experiments possible. I wish to thank them for the opportunity to visit their labs in Cambridge, MA and for their support of this. They made my stay over there most enjoyable and offered invaluable advice. Special thanks go to Joe Marks, Dr. Chia Shen, Clifton Forlines, Daniel Wigdor and Alan Esenther. I'd also like to thank Dr. Finian Buckley for his advice on the psychology aspect of this work, Dr. Michael Cooney, for his help on the statistics aspect of the thesis and Emilie Garnaud for the use of her Memory Game system.

To my friends in the CDVP, in particular Georgina and Pete. They have really kept me sane, especially during the write-up phase, and have always been there to ask questions, bounce ideas off, provide light-hearted relief during the low points and generally be great friends. I would also like to thank Cathal, Colum, James, Kirk, Mike, Neil, Paul and Sandra for the coffee breaks and chats that we have had and the help and advice that they have given me over the years. Similarly, to the rest of the CDVP members, both past and present, of whom there are too many to list, who have been a fantastic group to work with and have made my time doing this PhD very enjoyable indeed.

To all of my friends from home and from my undergrad years in college, especially Denise, Sheila, Rose-Anna, Mairéad, Carma, Nicola, Lisa and Ruth, thanks a million for being there for me, especially through the difficult times.

To Valerie, who helped me to regain focus when I felt I had lost my way during the tough times – thanks so much.

Special thanks go to my boyfriend Paul, who has been really understanding and supportive, especially over the past number of months, which has helped me greatly throughout this. Thanks for being there.

I have been truly blessed with the fantastic family that I have. To my brothers and sisters, Tom, Ed, Anne and Catherine, and their respective spouses and children, whom I want to thank for all of their support at all levels. They have been an inspiration to me and I am truly grateful for all they have done over the years.

Finally and most importantly, to my Mum and Dad, who have been the most fantastic parents and I thank them both from the bottom of my heart. They never ceased to be there to support me in all of my endeavours and look out for me whenever possible. My Mum is truly amazing and I'd especially like to thank her for all of her support over the years. It is with great sorrow that my Dad cannot share this achievement with me, but I know that he would be so proud right now. I have truly missed his presence over the past almost three years. It is to both my parents that I dedicate this body of work.

I gratefully acknowledge The Irish Research Council for Science, Engineering and Technology (IRCSET). Without its scholarship, this PhD would not have been possible. I would also like to acknowledge the support of Science Foundation Ireland under grant number 03/IN.3/I361, as well as the DCU Office of the Vice President for Research (OVPR) and KSpace for conference travel support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Multi-user and multi-touch technology is an area of computing that has been receiving much attention since its inception in the 1980s. The ability to have two people working on the same application in a collocated or a distributed setting, or to have one person using both hands to interact with an application, brings with it huge potential for developing exciting new applications that facilitate group-work, as well as novel single-user, multi-touch applications. In conjunction with the development of such applications, comes an urgent need to build these applications in such a way that they maximise the users' experiences of working with them, from both a functional and visual perspective. One approach is to look at the combined personalities of small groups of users working on a multi-user technology, to determine whether this has an impact on their interface preferences, performances and interaction. In this thesis, we focus on the design of applications for multi-user, tabletop technologies, taking into consideration the personalities of their potential users.

Developments in computing technology have been rapidly accelerating, in particular over the past 20 years (Teuscher and Hofstadter, 2004). Present-day computers are more powerful and pervasive than ever before, due to their high processing power, low cost and compact size. This has evoked the emergence of a number of new technologies, such as touch-sensitive devices and hand-held portable devices e.g. PDAs, laptops, mobile phone technologies etc. Now, even the combination of touch-sensitivity and portability has been made possible, highlighted by the recent release of Apple's iTouch and iPhone (Brewer and Bowcock, 2008). With these new technologies, arises a need to design applications that are appropriate and that visually and functionally meet the needs of their target audiences.

Most research in interface design has been catered to a single user working on a standalone PC application, though research into multi-user interface design is becoming increasingly popular. The commercial release of multi-user technologies such as Microsoft's Surface (Microsoft, 2007) means that addressing new application design issues for such technologies is now desirable. In the remainder of this chapter, we introduce the qualitative study, which we conducted to identify aspects of multi-user, tabletop application interfaces that pairs of users (dyads) liked or found useful. We do this with regards to the personality composition of the group members.

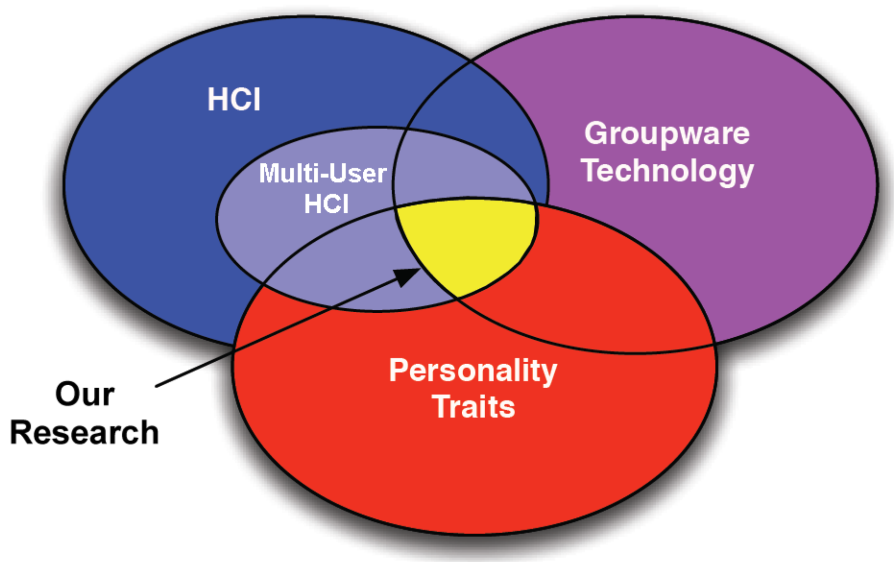


Figure 1.1: *Three research areas that our study is based on.*

1.1 Introduction to our Three Encompassing Research Areas

Here, we provide a brief description of the areas of Human-Computer Interaction (HCI), Groupware Technology and Personality Psychology (see Figure 1.1). One of the key contributions of this thesis is to provide a better understanding of the necessary elements of interfaces, such as layout and colours, and the factors affecting the design of interfaces that support multiple users (which we describe in greater detail in Chapter 2) that need to be considered. We do this with respect to the combined personalities of eighteen dyads, working on both collaborative and competitive collocated, multi-user systems.

1.1.1 HCI

"Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" (Hewett et al., 2008).

This includes the system and user requirements and specifications, design (both functional and interaction), testing, development and deployment. Figure 1.2 illustrates a general model of software design and development, which takes into account the users of the software application from the very beginning. These stages need to be constantly re-assessed as technology and user-needs change over time. From the outset, users are enlisted at the "Requirements and Specification" stage in order to elicit what exactly users want from the proposed software system, for instance, what features are essential in order for the system to be accepted and successfully accepted. Next, designs such as system interface mock-ups, system process documents (e.g. Data Flow Diagrams, UML) and architectural designs are proposed, some of which the user is shown (in particular interface mock-ups and proposed features) – hence, these designs are "tested" with real users. We can see from this diagram that the

"Design" and "Testing" stages are iterative, as issues arising from testing the system (both from a technical and user perspective) need to be fed back into the design of the system to remove inconsistencies and ensure that the system best meets the needs of its potential users. Once an acceptable design has been obtained, it is "implemented" and once again "tested" with users, another potentially iterative process. Finally, the system is deployed into the user's environment.

There are many other software lifecycle models available by which designers and programmers can design and build their applications e.g. agile software development, extreme programming or the waterfall model. However, it is outside the scope of this thesis to discuss these in-depth so we assume that the user-centric approach is the default one used.

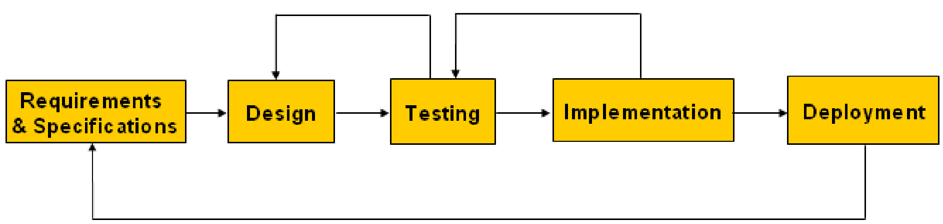


Figure 1.2: *Stages in system development that HCI is concerned with.*

Clearly, this is a very relevant area in this body of work, since it is concerned with the design of multi-user collaborative systems. However, more complicated issues are present for the design of multi-user systems than have been traditionally addressed in single-user HCI, such as user awareness, territoriality, division of labour etc. Hence, we focus on a special part of HCI, which we call Multi-user HCI. We describe this in more detail in Chapter 2.

1.1.2 Groupware Technology

Groupware systems are computer-based systems that support two or more users engaged in a common task, and that provide an interface to a shared environment (Ellis and Gibbs, 1989, p. 1). Groupware enables multiple users to work on the same application in various different settings, from distributed to collocated. The groupware technology that we used in our study is a multi-user, tabletop device called the DiamondTouch (Dietz and Leigh, 2001). We describe this technology in more detail in Chapter 3.

Examples of groupware technology include the popular and familiar e-mail and Instant Messaging (IM) systems, as well as less widely-known multi-user, touch-sensitive systems such as the The Lumisight Table and IPSI's Roomware. New applications are constantly being developed for these new technologies, while older applications, such as e-mail, are being upgraded to provide better facilities and support a wider range of user needs and wants. Again, this area is discussed in greater detail in the next chapter.

1.1.3 Personality Psychology

Personality psychology is one of the many sub-areas of Psychology (see Figure 1.3) and is mainly concerned with attempting to describe people in terms of their personal characteristics and their respective differences from other peoples'. In this thesis, we provide an overview of some of the more important and influential theories in Personality Psychology in an attempt to relate the personality of our experiment users with their performances and preferences on our multi-user systems.

Many theories and models have been developed in order to determine the personality types of individuals (e.g. the Myers-Briggs Type Indicator (Briggs Myers and Hauley McCauley, 1985), The Big Five (Costa and McCrae, 1993), Eysenck's PEN model (Eysenck, 1947)). The most successful and widely accepted model of personality among the Psychology community is "The Big

Figure 1.3: *Sub-areas of Psychology.*

Five" trait model of personality (Costa and McCrae, 1993), which is what we used to determine the personality profiles of our study participants. Generally for this, a person's personality profile is obtained using a questionnaire, which an individual must answer as truthfully as possible. We used the scores obtained along five personality traits from the results of this questionnaire, to compare to the performances and preferences of the study participants.

1.2 Objectives of our Research

With the development of multi-user, tabletop technologies, we have identified a research gap with respect to the design of applications for these technologies. Previous studies have been conducted on interface design, which take into account the personalities of individuals, but these studies were conducted for single-user interfaces (Reeves and Nass, 1996), (Karsvall, 2002), (Brinkman and Fine, 2005). These studies showed that personality had an impact on the preferences of users of the studied interfaces, as well as their performance in the case of (Reeves and Nass, 1996).

Building on such previous studies, the purpose of our research is to analyse the preferences, interactions and performances of dyads on different interface and constraint variants to five multi-user tabletop applications, which are both competitive and collaborative in nature. We then intend to determine whether these aspects are related in any way to the combined or individual personalities of the dyads (depending on whether the task is collaborative or competitive). As a result of this research, we can then advise on certain design issues, that should be taken into consideration when designing multi-user tabletop applications.

We also investigate the effect of issues, previously identified and studied in relation to multi-user systems (e.g. territoriality), on our tabletop systems and their respective interface and constraint variations. Some of the tasks we propose in this thesis are quite basic in nature, but since this is a very novel study, we decided to use these basic systems in our work in order to provide a foundation for potential future work on more complicated tabletop applications.

1.2.1 Thesis Hypotheses

In considering the purpose of our research, as described in the previous two paragraphs, we pose a number of questions which we aim to answer in this body of work, in order to determine whether a relationship exists between the personalities of dyads and their preferences and performances on multi-user technologies. If personality is found to have an effect, then this will have implications for the design of such multi-user systems e.g. the placement of objects on-screen, colours preferred. Specifically, the questions we pose are:

(1) Does the personality composition of a dyad (a group of two users) impact their performance on simple competitive and collaborative games and video search tasks?

What this proposes is that the personality types forming a dyad have a direct and measurable impact on their performance of the task at hand. This means that we should be able to measure a difference in performance and correlate that in some way with the personality composition of the dyad.

(2) Do dyads prefer certain interfaces and perform better on these interfaces?

This proposes that certain combinations of personality types within a dyad will show a marked preference for certain interface characteristics and that they will perform a specific task better on an interface that they prefer.

and

(3) Is the personality composition of dyads related to the manner in which these dyads carry out such simple competitive and collaborative games and video search tasks?

This implies that personality combinations making up dyads, affect the interaction as well as performance of the dyads in collaborative and competitive tasks.

From these research questions, we formulated our three main thesis hypotheses, which we aim to prove or disprove as a result of the execution and analysis of our study. Further, we broke these overall questions into a number of hypothesis sub-questions, which by posing, would to support or disprove the overall hypotheses. These are hypotheses and respective hypothesis sub-questions are:

Hypothesis 1: The personality composition of a dyad impacts the performance of that task, or in other words, dyads composed of certain personality types will perform tasks better than others.

Q 1. Do we simply focus on Extraversion as the sole personality factor to correlate to performance or interaction style of dyads ?

Q 2. Do the remaining “Big Five” personality traits affect the performances of dyads ?

Q 3. Do dyads that are more similar in terms of their personality composition outperform dyads containing very different personality types ?

Q 4. Is the interaction recorded among dyads related to their personality composition ?

Hypothesis 2: Dyads with certain personality types will prefer and work better on certain interfaces.

Q 5. Do individuals within dyads develop a similar impression of a system ?

Q 6. Do users prefer interfaces which model their personality along the *Extraversion* trait ?

Q 7. Do dyads perform better on an interface variant/under a task constraint variant that they like better when give two variants ?

Q 8. Is there a relationship between a user’s stated opinions on a system and their interaction data ?

Hypothesis 3: Dyads perform different tasks in a different manner and this is related to their personality.

Q 9. How does imposing different constraints on a collaborative task affect the performances of the dyads ?

Q 10. Are there more interaction instances in a collaborative version of a game as opposed to a competitive version ?

Q 11. Does the amount of interaction among a dyad relate to the performance of that dyad in our collaborative tasks ?

Q 12. Do dyads coordinate their actions well on our collaborative search tasks and is this related to their personality type ?

Q 13. Do the same territoriality tendencies exist regardless of the task or are there cases of some tasks where territoriality is irrelevant in both our competitive and collaborative tasks ?

Q 14. Do dyads with certain personalities employ different territoriality techniques than others when performing all of our tasks ?

We also posed two other questions to investigate and answer, which are related to the validity of our data. Specifically we ask:

Q 15. Does performance of dyads vary to a greater or lesser extent across the different collaborative tasks used ?

Q 16. How much variability is there in the interaction among dyads across the different collaborative tasks used ?

1.3 Thesis Structure

The layout of the remainder of this thesis is as follows. Chapter 2 provides us with an description of the three key research areas that this body of work encompasses – those being Human Computer Interaction, Groupware Technology and Personality Psychology. Included in this explanation is a literary review of related work in each of these three areas, providing many of the foundations on which our experimental systems and subsequent analysis were built. In Chapter 3, we provide a background of our own work undertaken prior to this study. This includes a description of a digital video search system that we designed, developed and implemented on a multi-user collaborative tabletop device, called the DiamondTouch (Dietz and Leigh, 2001). We also describe each of the systems that we built for our user-experiments, which we designed based on the previous research discussed in Chapter 2. In the final part of this chapter, we assert the hypotheses that we hope to prove or disprove as a result of our user-experimentation.

Chapter 4 describes our experiment methodology, including a description of each of the systems that we used and their respective interfaces or constraint variants. We also include a description of how the participants were recruited and selected to carry out our user-experiments and what they were required to do as part of these experiments. In Chapter 5, we give a detailed analysis of the results of our experiments, including a description of the data we gathered and how certain elements of this data related to other elements. From answering a number of sub-hypothesis questions based on this analysis, we can then prove or disprove the hypotheses of this thesis. We conclude this thesis in Chapter 6 by summarising what we have learned from our user-experiments and subsequent results, the guidelines we can draw from this knowledge in terms of designing the interfaces to multi-user tabletop applications and the possible future work that could be undertaken to further explore and solidify the phenomena presented.

Chapter 2

Background and Related Research

In this chapter, we provide a detailed background and literary review of the three areas that this body of work incorporates. These three areas are Human Computer Interaction (HCI), both in a single-user and multi-user context; Groupware Technology and the different categories that it comprises; and finally Personality Psychology and some of its most popular theories and tools, as well as its significance in the media and in the area of computing.

2.1 Human-Computer Interaction

HCI, also known as Computer-Human Interaction (CHI) or Man-Machine Interaction (MMI) is:

> "concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena around them." (Hewett et al., 2008, p.1).

The field of Human-Computer Interaction (HCI) is an ever-growing area of research, which fuses expertise from numerous different disciplines in order to design and create efficient, effective and satisfactory software systems and interfaces. These disciplines include Computing, Engineering, Anthropology and Behavioural Science, which take into consideration both human aspects of what should be provided for and the technical aspects of what can actually be provided when designing a software system. Research to date in HCI and usability has largely focused on single users working on separate computing devices (e.g. Nielsen, 1993; Nielsen and Mack, 1994; Doan et al., 1995; Shneiderman and Plaisant, 2005) while research in design for multi-user devices is still only emerging.

Increasingly, organisations and individuals are realising the importance of HCI in the development of their software products (Nielsen, 1993). Highly competitive markets compel system developers to allocate a substantial amount of effort to designing the user interface, to ensure that users can carry out their system-supported daily tasks in an efficient, effective and enjoyable manner.

A system could carry out all of the desired user-functions, but if the interface to it is not intuitive, is difficult to learn and causes obscurely worded and unexpected error messages, it will inevitably be rejected by its users. A poorly designed interface can damage a system developer's reputation affecting its current and future software sales and acceptance (see Figure 2.1). Hence, investing considerable time and effort into designing an interface to meet the needs of a system's target audience is necessary to maximise the likelihood of its success.

In this section, we describe aspects of HCI research that have been undertaken with regard to single-user environments, including principles and guidelines[1] that can be utilised in and extended for use in multi-user, collocated environments (Nielsen, 1993; Nielsen and Mack, 1994; Norman, 1998; Shneiderman and Plaisant, 2005). We then describe research into the design of multi-user systems and also include sections on specific design considerations for such sys-

[1]Here, a guideline refers to "A statement or other indication of policy or procedure by which to determine a course of action"(American-Heritage-Dictionary, 2008).

Figure 2.1: *Extreme user frustration at a "bad" system interface (Smilemania, 2008)*

tems, including widget placement, awareness, coordination policy and division of labour. The next subsection provides an overview of HCI research that has been conducted for single-user interfaces, some of which can be extended for use in multi-user, collaborative computing devices.

2.1.1 Single-User HCI

Computing devices are used for a wide range of purposes; examples include as work support tools, and information sharing, communication and entertainment devices. These range from devices in mobile form (e.g. mobile phones, laptop computers) to desktop PCs and mainframes. Rapidly evolving technologies mean that computers are continually becoming faster and more powerful, with almost limitless storage availability. Since computers are such an integral part of the everyday life of modern society and are fast evolving, designers need to constantly re-evaluate and if necessary, revise their applications and respective interfaces to cater for more diverse audiences, for new and improved technologies, and in order to keep pace with competition. In addition, these demands constantly challenge system and interface designers to expand and integrate new revisions to systems in a natural and intuitive way.

Figure 2.2: *Single-user, single-PC HCI*

The demand for computers has been further accelerated due to the widespread popularity and use of the Internet and World Wide Web. The uses and possibilities of computers are endless. But what about the applications developed for these computers, that people interact with in order to accomplish their goals/tasks? In the early days of computing, interface design relied primarily on the system designer's own intuition and experience. However, application developers soon realised the importance of providing interfaces to their systems that met the needs, wants and expectations of the people who would use them. An application interface that does not satisfy the needs and expectations of its' targeted end-users is at the very least disappointing, and more often than not, unusable.

Terms such as User-Centred Design, Usability and User Experience have been the focus of much research and analysis in the field of HCI. "Know the user", Hansen's famous first principle of user-engineering, underlines the importance of injecting users' wants, needs, concerns and opinions into the design process (Hansen, 1971). Understanding the target audience of a software product and becoming aware of the routines and daily processes of that target audience is essential to the system's acceptance, and long-term and widespread deployment.

Guidelines and Principles

Numerous sets of guidelines, principles and theories for interface design have been devised. Most of these are the result of designers' experiences, common sense and the application of such 'design principles' that many HCI-text books deal with. One example is Nielsen's (Nielsen and Mack, 1994) 10 Usability Heuristics, which are as follows:

1. **Visibility of system status:** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

2. **Match between system and the real world:** The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

3. **User control and freedom:** Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

4. **Consistency and standards:** Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

5. **Error prevention:** Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

6. **Recognition rather than recall:** Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another.

Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. **Flexibility and efficiency of use:** Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

8. **Aesthetic and minimalist design:** Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

9. **Help users recognize, diagnose, and recover from errors:** Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

10. **Help and documentation:** Any help and documentation should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Shneiderman also developed a set of interface guidelines (Shneiderman and Plaisant, 2005) known as his eight golden rules of interface design and these are:

1. Strive for consistency.

2. Cater to universal usability.

3. Offer informative feedback.

4. Design dialogs to yield closure.

5. Prevent errors.

6. Permit easy reversal of actions.

7. Support internal locus of control.

8. Reduce short-term memory load.

Norman also formulated a number of general interface design principles which he developed from his insights into the field of industrial product design (Norman, 1998). These include:

1. **Visibility** - make functions visible.

2. **Feedback** - audio, tactile, verbal, visual, and combination.

3. **Constraints** e.g. deactivate certain menu items.

4. **Consistency** - easier to learn and use. Difficult to maintain for more complex interfaces.

5. **Affordance** - e.g. graphical elements like buttons, icons, links, and scrollbars, which make it appear obvious how they should be used.

These all provide excellent, generalised advice on what should and should not be done when designing an interface. One notices that these guidelines are not specific to any particular device, platform or modality. More device-specific guidelines are available e.g. for mobile phone devices

One of the key contributions of this thesis is to provide a better understanding of the necessary elements, such as layout and colours, and the factors affecting the design of interfaces that support multiple users (which we describe in greater detail in Section 2.1.2 below) that need to be considered. We do this with respect to the combined personalities of eighteen dyads, working on both collaborative and competitive collocated, multi-user systems.

Knowing the User and their Surroundings

In addition to design considerations for the software application itself, such as the guidelines, rules and principles mentioned in the previous section above, there are also physical considerations that need to be taken into account, i.e. ergonomic concerns. These include chair and monitor height and the ability to adjust these, keyboard designs, reach distance. One of the most important models, which relates to reach distance is Fitt's Law (1994). This model predicts the time required to move quickly to a target area, with regards to pointing, both in real-world and computer terms e.g. mouse pointing. It has been very successful at predicting such times and has been applied to tasks where a user working on a PC Graphical User Interface (GUI) must position the mouse cursor over a target on the screen e.g. a widget. Both point-and-click and drag-and-drop actions can be modeled by Fitt's Law. However, it only applies to movement in a single dimension, describes simple motor reactions of, for example, the human hand and describes movements that are untrained or practiced.

The psychology community have also modeled user behaviour in terms of their reactions to computer interfaces. In their book "The Psychology of Human-Computer Interaction" (Card et al., 1983), the authors described a Model Human Processor, in which a user's interaction with a computer could be divided into three interacting subsystems: (1) the perceptual system, (2) the motor system and (3) the cognitive system, each with its own memories and processors. They stated that the perceptual system consists of sensors and associated buffer memories, the most important buffer memories being a Visual Image Store and an Auditory Image Store to hold the output of the sensory system while it is being symbolically coded. The motor system translates thoughts into actions, by initiating patterns of voluntary muscles, which are arranged in pairs of "agonists" and "antagonists", executed one shortly after the other (Card et al., 1983, p. 34). The arm-hand-finger system and the head-eye system are the most relevant and important sets of effectors for computer users.

The cognitive system receives symbolically coded information from the sensory image stores in its Working Memory and uses previously stored information in the Long-Term Memory to make decisions about how to respond. Modeling users in this way, provides a common framework in which models of memory, problem solving, perception, and behaviour can all be integrated with one another. This can in turn be used in order to decide what menu items, widgets and functions should be present in the interface, how they should be represented and where they should be placed.

Sharp and colleagues (2007) noted that there were multiple aspects of the user experience that could be borne in mind and ways of considering them in the design of interactive products. Of pivotal importance are the usability, functionality, aesthetics, content, the look and feel, the sensual and emotional appeal of applications, as well as taking into account the sensitivities of the application users. The phrase *User-Centred Design* involves including users at every stage of the design process. This includes observing users carrying out their daily tasks and functions, talking to them, interviewing them, asking them to carry out performance tasks, modeling their performance, asking them to complete questionnaires, and even asking them to become co-designers.

From all this we can see that becoming very familiar with our users is vital. It is important that users' personalities, skills, experience, background, age, ethnicity and position in an organisation be known and catered for. For example, a system that potentially threatens someone's position and esteem in an organisation will cause discomfort and annoyance and may cause the software to be rejected, regardless of its utility and effectiveness. In addition, designing for novice, knowledgeable and expert users all in the same interface can also be a challenge, but is important to ensure the satisfaction and system acceptance of users at all levels.

The concepts and guidelines mentioned are in most cases general enough to apply to both single-user and multi-user devices alike. However, multi-user interface design brings with it added complexities unseen before in single-user HCI. This makes creating guidelines for such multi-user systems a much more intricate task. Many of these complex factors are discussed in Section 2.1.2, where the various aspects of multi-user HCI, such as territoriality and orientation are discussed. In this thesis, we look at the combination of users working on multi-user collocated systems and how the opinions and personalities of these users affect their interface preferences and performances on such systems.

Evaluating Interactive Interfaces

Usability measures, usability tests and review processes (Nielsen, 1993; Shneiderman and Plaisant, 2005) have been formulated for evaluating interfaces at various stages in their development. Traditionally, information retrieval systems, for example, have been evaluated for their 'effectiveness' in terms of Precision and Recall. These two measures have been used to quantify the search engine's back-end performance (Van Rijsbergen, 1979). However, as the interactivity of such systems has become more important, how to evaluate the user interaction has also become an important research agenda in HCI. This is also true of information retrieval (IR) systems, the evaluation[2] of which can take the form of one of the following three evaluation methods[3] as discussed in (Borlund, 2003):

1. The system-oriented approach (i.e. the Cranfield (Cleverdon et al., 1966; Cleverdon and Keen, 1966) model of evaluation, which involves the use of test collections, also (Robertson and Hancock-Beaulieu, 1992))

2. The user-oriented approach (i.e. views the Information Retrieval environment as dynamic and relevance as individually and situationally based (Lancaster, 1969; Saracevic and Kantor, 1988a; Saracevic and Kantor, 1988b))

3. The hybrid/combined approach (i.e. which combines the experimental control aspect of the system-oriented approach with the dynamic and individual nature of information needs Borlund, 2003)

[2]Evaluation here is "The process of examining a system or system component to determine the extent to which specified properties are present" (Dictionary-Of-Computing, 2008).

[3]Here, we define methods as "The procedures and techniques characteristic of a particular discipline or field of knowledge" (OED-Online, 2008).

Each of these methods have advantages and disadvantages, which are discussed in (Sparck Jones and Willett, 1997). The main differences between the first and second methods are that information needs are viewed as static and variables are kept controlled in the first, whereas information needs are viewed as dynamic over time and individually-based in the second. Another example of a system-oriented approach to IR evaluation is the TRECVid conference series, which is discussed in greater detail in Chapter 3. The main advantages of this type of approach are:

1. It offers an "unrivalled series of direct performance comparisons of retrieval techniques" (Sparck Jones and Willett, 1997).

2. Its test collection is limited in size (although the collection is still large) and pooled relevance judgements are available, so that retrieval performance is easy to measure.

3. It operates a set of well-defined measures of success in terms of success in retrieving relevant documents i.e. precision and recall.

4. Research using test collections identifies good retrieval technology, allowing expensive user testing to be reserved for the most promising avenues, for instance basic components of current web search engines were initially developed for test collections (Voorhees and Harman, 2003).

The main disadvantages are:

1. Statistical significance tests in TREC are weak since the population distributions underlying the observed performance values are not known, however the use of significance tests among participants is increasingly being encouraged.

2. It is typically concerned with success in retrieving relevant documents only.

3. It fails to engage substantively with the evaluation of online interactive searching, especially by end users and with other environment variables.

4. Precision is calculated over large document sets, more than a user would ever consider (our approach was to consider recall) and when averaged, can be too summary in nature to be effective.

The main advantages of the user-oriented approach are:

1. Statistical significance tests can be used to determine the significance of the output performance of the system itself and establish their correlation with the different environment variables.

2. It concerns more than just retrieval of relevant documents, but other variables in terms of the user's interpretation of relevance, their constantly changing information needs etc.

3. Metrics other than Precision and recall are used to determine system effectiveness.

The main disadvantages of the user-oriented approach are:

1. Evaluation methodologies are still evolving and do not beat TREC's direct performance comparisons of retrieval techniques.

In (Borlund, 2003), the author proposed a third approach which combined both methods 1 and 2 (i.e. the collection of system-oriented data and cognitive user data). Here a simulated situation is used, which comprises of a simulated work task situation and an indicative request.

In his book *Usability Engineering*, Nielsen (1993) devised a list of methods for evaluating the usability of interactive interfaces. These are:

1. Heuristic Evaluation

2. Performance Measures
3. Thinking Aloud
4. Observation
5. Questionnaires
6. Interviews
7. Focus Groups
8. Logging Active Use
9. User Feedback

Often these evaluation tests, inspections and methodologies are implemented prior to the release of a system in order to catch as many errors, flaws and inconsistencies as possible. Any errors caught after this phase are comparatively difficult and expensive to fix. Similar to evaluating information retrieval systems using Precision and Recall, there is usually a set of "measures" that can be obtained from conducting these types of evaluation. For example, the generally agreed five usability criteria devised by (Nielsen and Mack, 1994) are used as concrete measures:

1. **Efficiency** (time taken to complete a task)
2. **Learnability** (time taken to become an expert user)
3. **Memorability** (retention rate over time)
4. **Error rate** (number of times the user pressed wrong button)
5. **Subjective satisfaction** (questionnaire rating scale 1-5)

While the above criteria and their corresponding measures are useful for desktop computer interfaces, where an individual uses the system, any other

Types of Usability Tests	
Paper mockups:	Images of screen displays given to users to get their opinions.
Discount Usability Testing:	Testing an interface on a small number of users (3 - 6 people).
Competitive Usability Testing:	Comparing a new interface to previous versions of the same interface, or to the interfaces of competitors.
Universal Usability Testing:	Testing a new interface on a wide range of platforms and devices and on a diverse range of users.
Field Tests and Portable Labs:	Placing a new interface in a real-world environment for a trial period of time.
Remote Usability Tests:	Online usability tests.
Can you Break This Tests:	These are often given by game designers to see if users can beat the game.

Table 2.1: *Types of Usability Tests*

context that deviates from this particular setting will require different, more tailored criteria as well as suitable measures for them. For example, we believe that evaluating a mobile chatting service implemented on a PDA would benefit by having an extra criterion such as:

6. **Mobility** (time taken to get back to interaction after looking away)

Similarly, for evaluating a multi-user tabletop interface, such as used in this thesis, we believe it would be beneficial to have an extra criterion such as:

7. **Collaboration** (how well the interface supports interaction between the co-users)

A number of usability tests can be carried out at various stages throughout the development of an interface, as listed in Table 2.1 (Shneiderman and Plaisant, 2005). Determining which of these usability tests is most appropriate for evaluating a software system is a decision the designer must make in order to give full and adequate feedback on all aspects of the system and its interface.

User surveys with focused items that achieve specific aims set prior to their dispersion, along with preselected statistics are also very beneficial. In testing the reactions of users to a word processor, Coleman and Williges developed a survey containing a set of bipolar, semantically anchored items, which proved to be very effective (Coleman and Willeges, 1985). For tabletop applications, interface evaluation would benefit too from such directed surveys to reveal users' reactions e.g. whether a pair of users develops a similar impression of the system after working with each other or not and what causes differing opinions.

In investigating multi-user tabletop interface design in this thesis, we attempt to explore what criteria exist in successful application design, by carrying out a series of user-tests on collaborative and competitive systems and how these criteria can help reasonably evaluate a tabletop interface.

2.1.2 Multi-User HCI

The advent of groupware technologies (see Section 2.2), supporting a wide range of multi-user tasks/groupwork has brought with it an emerging focus on multi-user HCI research. In conducting research into multi-user interaction design, it is not just a single user's interaction with a groupware device that must be considered. In a groupware situation, considerations must be taken by designers for three main factors:

1. The interaction between each of the members of the group
2. The interaction between each individual user and the technology
3. The interaction of all users with the technology and the respective issues that this presents.

These factors can be visualised, as in Figure 2.3.

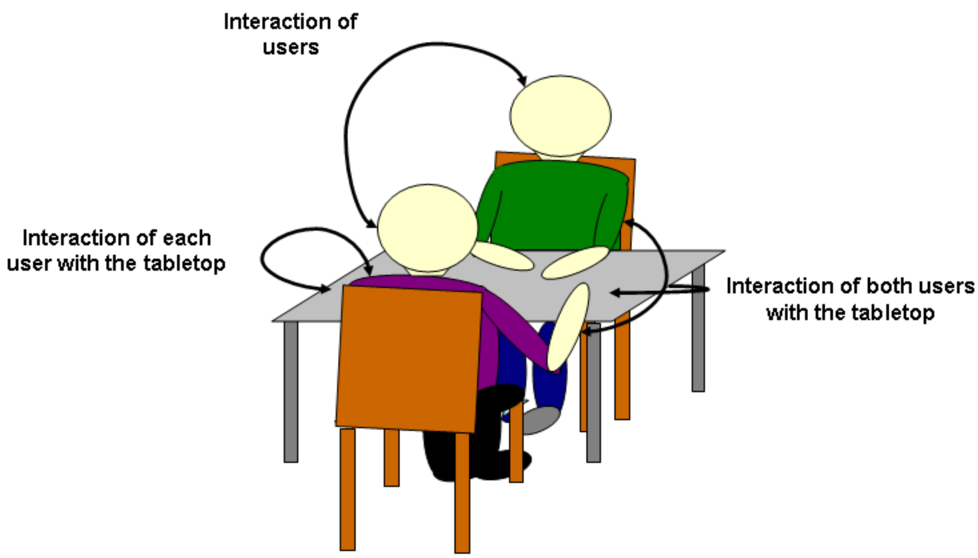


Figure 2.3: *Designing for a multi-user tabletop device*

Thus HCI research for groupware brings with it a complexity unknown in traditional HCI. Much research has been carried out in providing the groundwork for building successful groupware products, through observing groups as they use prototype or experimental systems (e.g. Ellis et al., 1991; Gutwin et al., 1996; Ringel Morris et al., 2004b; Ryall et al., 2004; Ringel Morris et al., 2006). Phenomena such as territoriality, widget placement and awareness have all been studied and strong trends observed, thus improving our knowledge about what works, what does not work and what is desirable in groupware system design. However, no research has yet been conducted into the relationship, if any, between the personalities of the members of a group and their preferences, performance and interaction in the design of interfaces to groupware applications. This gap is where our research focus has taken us, though specifically for the design of interfaces to collocated tabletop systems.

In the following sub-sections, we discuss research studies that have examined territoriality and orientation, awareness, widget placement, division of labour and coordination policy for groupware systems. All of these topics have a major influence on our work in optimising interfaces to each group of two users, based on the personalities of the people in that group. We also include a brief section on various modalities available for use in multi-user applications.

Territoriality and Orientation

Observing what parts of an interface each individual in a group works on for various parts of a task is important for multi-user software design. Studies by (Scott et al., 2004) found that individuals used three areas when working collaboratively on a tabletop: personal, group and storage areas. Each of these areas were defined by the social protocols and norms of the subjects studied. By social protocols here, we mean the social skills that we as individuals, have developed and evolved such as politeness and etiquette. They found that the table edge directly in front of a user was utilised as that user's personal territory, in which they could perform their own independent tasks. Objects located in personal territories were oriented to the person "owning" that territory.

Group territory was found to occupy all areas of the tabletop surface, with the exception of the personal territories established in front of group members. Objects in group territories were oriented in a direction that was most visible to all group members. Users were found to execute most of their collaborative work in group areas that were located closest to them. Storage territories were used to store task resources in piles (e.g. groups of pictures piled on top of each other) and such territories were found to be located anywhere that the groups were not working, i.e. there were no specific storage areas - they were mobile depending on the group's work strategy. In their study[4], storage territories existed on top of personal and group territories. They also found that each of these tabletop territories had functional and spatial properties.

These findings have important implications for designing tabletop interfaces and also link in with the division of labour policy selected (see Section 2.1.2) i.e. if groups work in series, in parallel, or a combination of both.

A study into the effects of table and group size in collocated settings has also been carried out. In (Ryall et al., 2004), the authors found that while groups reported an overall preference for a larger tabletop, the table size did not affect their task efficiency or distribution of work. However, the size of the group itself affected the speed at which the given task was completed.

Reach distance was examined by (Toney and Thomas, 2006). For instance, regions of over-lapping reach of group members was found to relate to the group territory described above, and table sizes were found to directly affect the sizes of these territories, i.e. the bigger the size of the table, the smaller the reach-overlap region and hence the smaller the group territory size (Scott et al., 2004).

Orientation is a phenomenon that is strongly related to territoriality. An object that is oriented to an individual is not only more understandable by that individual, but also communicates to others in a group that that particular individual is using the object and it is not publicly available for use by the others. (Kruger et al., 2003), noted that orientation was not just used for individual comprehension, but that it had a range of other purposes. These were *comprehension* (ease of reading, ease of task and alternate perspective), *coordination* (establishment of personal spaces, establishment of group spaces and ownership of objects) and *communication* (intentional communication and independence of orientation). From a communication perspective, orienting an object to another user/group of users was a very direct way of establishing an

[4]A study here refers to “A study here refers to ”a careful examination or analysis of a phenomenon, development, or question” (Merriam-Webstar, 2008).

audience for what that person wanted to say about that object.

From a distributed perspective, the findings of Scott and colleagues' study have also been supported for a distributed tabletop setting (Scott et al., 2004). In (Tuddenham, 2007), a study was conducted on distributed tabletop territoriality and orientation. Each participant had a rotated view of the surface and could see other distributed group members actions as they were being executed. The fact that their view was rotated in relation to other users meant that the other members of the group were virtually at the other sides of the table, and objects were placed in personal territories to these other users and oriented to them. He found that the territories discovered in (Scott et al., 2004), held for this distributed setting, as did the orientation findings from Kruger's (2003) study.

The discovery of these various aspects of territoriality and orientation have a huge impact on design for both collocated and distributed multi-user technologies. Generally, using non-personal territories to hold group controls and display areas is advisable. Controls that support individual work can be located in each user's personal territory. Also, the ability to orient objects to a direction desired by a user is of vital importance in supporting these three intricate facets of collaboration. All of these facets have been addressed in the systems we have used in our study.

Awareness

A subject of much interest in the study of multi-user HCI is the issue of awareness. Numerous lines of research have found that a user's awareness of the other members of their group, where they are working, what they are doing, what objects they are manipulating and what their intentions are, is essential to supporting and ensuring the success and personal satisfaction of groupware-supported activities (Gutwin et al., 1996; Gutwin and Greenberg, 1998; Smeaton et al., 2006a; Villa et al., 2008).

In physical face-to-face environments such as a group working with real-world artifacts over a table, awareness does not need to be enhanced. Users can see what others are doing and objects can only be manipulated in a way that is visible to others. This is not so with virtual objects in groupware, particularly in the case of distributed, synchronous groupware. Awareness-providing widgets and cues are vital to users in such a virtual environment, to ensure that group collaboration and coordination is facilitated and successful in a not-so-natural environment.

These issues are skillfully identified and dealt with by Gutwin and colleagues in a number of conference articles. In (Gutwin et al., 1996), they tested a number of awareness widgets in a distributed, synchronous, groupware-supported construction task. The task required nine pairs of users to layout newspaper content across two pages. The users worked on two different PCs located opposite each other, with a partition between them. Their shared workspace in this case was the two pages. Additional widgets were supplied to different pairs, that increased each user's awareness of the other person's location of work on the workspace and the objects that they were currently working on. The study identified two awareness widgets that participants found particularly useful - those being radar views and mini views.

Mini views were boxes located at the top left-hand corner of the user's computer screen, which showed the entire workspace at a scale 64-times smaller than the actual workspace. Each of the objects on the workspace were represented by rectangles of different colours. Users could see some changes issued by their partner, as the movement of objects was mirrored in this view. Radar views were similar to mini views, though with the radar view, each user could see where their partner was working, through telepointers that represented each participant's mouse cursor. Each participant could be identified by displaying their telepointers in a colour that was associated with that user, as well as showing an outline of their main view. WYSIWID (what you see is what I do) views and multiple scroll-bar views were not deemed very useful for this task.

The authors also found that participant-supplied awareness information, for example, one user telling the other what object they were working on, was evident in this study. This helped prevent their partner from working on the same object. Hence, from this we can determine that awareness information with regards to what other people in a group are working on is essential to the successful coordination of actions and tasks and removes any aggravation or frustration caused by two users attempting to manipulate the same object. Users in this study also suggested that information be supplied about each user's intended work, for example, tagging objects that a user intends to work with next. Another suggestion was enable users to click on their mini view of their partner's main view, to directly interact with the objects others are working on e.g. in order to help them. In other words "improve the bridge between perception and action".

A later publication suggested a number of solutions to mitigate the tradeoff between individual power and workspace awareness in three areas of groupware design (Gutwin and Greenberg, 1998). These areas are workspace navigation, artifact manipulation and view representation. They suggested radar and detail views, over-the-shoulder views and cursor's eye views as means of maintaining awareness of where other users are working, the objects they are manipulating and what functions they are invoking on these objects.

The authors also made a number of suggestions for heightening user awareness of others' actions. For example, artifact manipulation is concerned with improving users' awareness of symbolic command invocation (which do not have a physical counterpart). These type of commands involve using menus, buttons, keypresses etc. and are difficult for other users to see. Techniques suggested for making these commands more visible to other users included remotely visible pop-up menus, group-visible dialog boxes, symbolic indicators and "supernova" animated key actions, such as the "delete" action invoked by a keypress. In the study, deictic and gestural communication was lost when using different representations of the same workspace. This was solvable by using fisheye views with different levels of distortion or different font sizes, through a mathematical transformation.

Villa and colleagues (2008) conducted a study on the impact of awareness on a synchronous, competitive search task for groups of two users, working on separate PCs. The authors defined four conditions - one where each user was aware of what their opponent was doing (i.e. they could see their opponent's display), one where one user was aware of the actions of their opponent (a watching condition), one where a user was aware that their opponent could see their actions (a watched condition) and finally, a condition where both users worked independently. The results here found that performance did not necessarily improve in the awareness conditions, but providing awareness information did reduce the amount of searching and the degree of effort for each user.

Since the systems we designed and built for this body of work are for a collocated setting, specifically a multi-user tabletop device, awareness is not quite as complex an issue here as for distributed environments described above. In a collocated setting, peripheral vision enables each user to see what the other is doing at any given time. However, if users are working intently on a sub-task they are doing individually at opposite sides of the table, awareness cues should explicitly be provided, such as sounds when a particular function is invoked or "supernova" animations when key actions are invoked.

Figure 2.4 illustrates how increasing levels of awareness can be provided depending on whether the design is for a distributed multi-user setting or a collocated setting. It can be seen from this that the awareness cues in the distributed setting are much more fundamental, as this type of setting naturally provides very little awareness information. Awareness cues in the collocated setting enhance and complement the strong natural user-awareness that is already present.

In (Smeaton et al., 2007), we report on experiments[5] that we conducted for the TRECVid benchmarking conference series, which compared two versions

[5] An experiment here means "An action or operation undertaken in order to discover some-

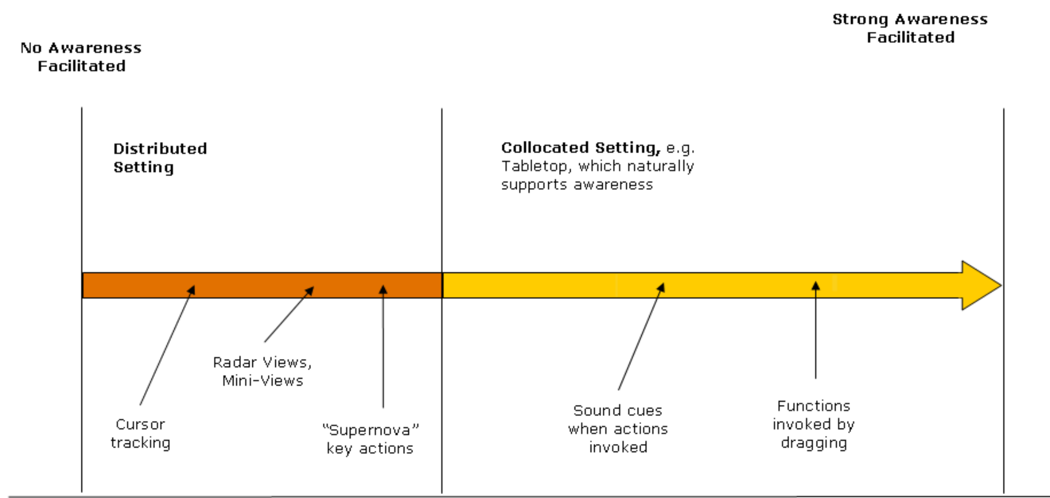


Figure 2.4: *Awareness techniques*(Smeaton et al., 2006a)

of a collocated, collaborative video information retrieval system with different levels of awareness. These different levels of awareness were designed and implemented based on Figure 2.4, where one interface had all of the awareness facets listed for the collocated setting, while the other interface did not. The study found that in general, most of the test participants preferred the interface that provided the most awareness. This was because they found the double-tapping on the lower awareness interface more tedious, the dragging metaphor used on the awareness interface was more natural and intuitive, the awareness interface enforced greater cooperation and coordination, was more convenient and was easier to get accustomed to. Hence, we selected this interface for use in our main body of experiments (see Chapter 3 for further details).

thing unknown, to test a hypothesis, or establish or illustrate some known truth" (OED-Online, 2008).

Widget Placement

An issue that is raised in designing any interactive system, is where widgets such as buttons, menus, text-boxes etc should be placed. In (Ringel Morris et al., 2006), the authors determined that the majority of people in each of six groups (each group comprising four people) preferred replicated widgets/controls to centralised ones. The reasons for this were that users disliked touching controls within .5 seconds of another user (co-touching) in case they collided, and centralised widgets took up important screen space in the centre of the table, an area that users felt was important for accomplishing shared tasks or carrying out group-work.

With regard to our research, we were curious as to whether expressed user-preference in widget placement had any relationship to the personalities of users. We suspected that users who were *extraverted* in nature would prefer shared widgets/controls as this caused them to work in a more sociable manner, whereas *introverted* people would prefer being able to work quietly and individually with duplicated widgets. This was one of the questions we posed in the study of our fifth tabletop system (see Chapter 4).

Division of Labour

Once given a task, groups need to decide how to break up the task among the group members e.g. should they work together constantly, should they split the task and work independently or should they work together for a time and then work independently on a subset of tasks. This is an important consideration in the design of a multi-user interface, as the interface must support the type of task-division chosen by its group members.

Figure 2.5 illustrates a number of ways that a task can be divided among a dyad (a group of two users). The first depicts one user doing their part of the task and passing their work onto the other user, who completes the task. The

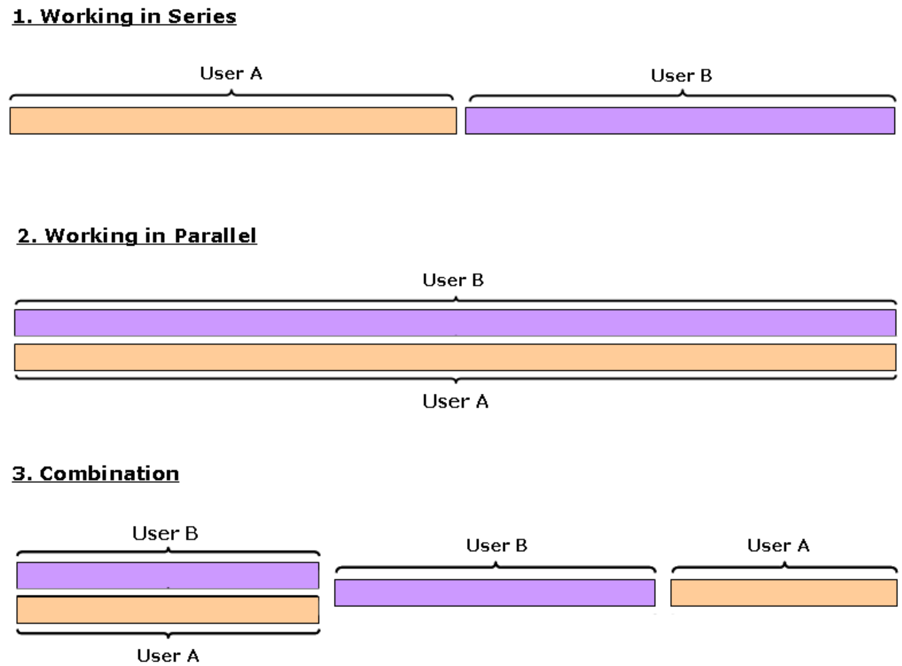


Figure 2.5: *Ways of dividing a task for a dyad*

second scheme shows both users working independently to achieve the end goal. The third exemplifies a hybrid collaborative effort, where both users conduct some subtasks in parallel, and then work with the results of these subtasks in series.

Coordination Policy

Should a multi-user application depend on social protocols to support a natural coordination policy, or should it explicitly enforce a particular coordination policy by applying software-level constraints to particular members of the group? In their study on spatial separation and partitioning in single display groupware, (Tse et al., 2004), observed that two people working together on a PC with two mice worked on the parts of a drawing and tracing application that were closest to them spatially. They also used social protocols to avoid interfering/hindering each other's actions. Studies prior to this created interaction techniques that facilitated the enforcement of social protocols at a software level, an example of which is found in a study conducted by (Ringel Morris et al., 2004b). Tse and colleagues' study proved that this was not absolutely necessary, thus changing the focus from building interaction techniques, to supporting a group's natural behaviour.

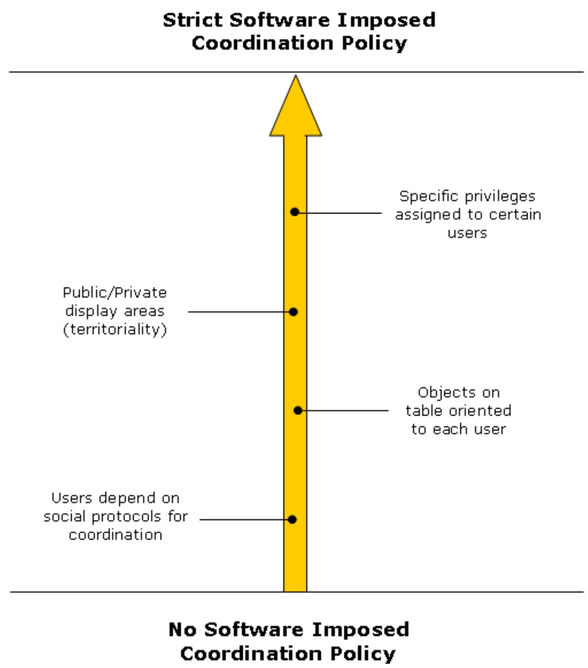


Figure 2.6: *Levels of software-imposed coordination policy*

Issues still arise in relation to data privacy and the privileges (e.g. access privileges, where only certain users with particular roles are allowed to read and/or modify certain objects, documents etc.) that some users have in terms of modifying objects on a collocated tabletop device. In such cases, software protocols may be enforced on users to explicitly prevent them from modifying objects that they should not modify. Whether to support such protocols in a groupware product or not is a decision that the designer and recipients of the software must make.

Figure 2.6 (Smeaton et al., 2006a) conveys some ways in which software can enforce coordination to different degrees. At the lowest level, the application relies entirely on the social protocols exercised by the group members themselves. Increasing levels of software-imposed coordination policy are provided by implementing orientation techniques, specifying public and private areas and assigning access or modification privileges to certain users.

Multi-User HCI Modalities

Another consideration for multi-user systems is the different modalities that can be used. In (Tse et al., 2006), developed an engine that enabled both speech and gestural input into two different Geospatial applications built for use on the DiamondTouch (Dietz and Leigh, 2001) tabletop device. Both forms of input complemented each other and allow for a richer user experience. However, issues with regards to users working in different modes, larger group sizes, participants with different roles and conflict situations had still to be tackled.

In (Ringel Morris et al., 2004a), the authors created a system that handled public and private audio output, along with gestural input. The purpose of the audio channels was to realise the notion of Single Display Privacyware, whereby private information could be distributed to specific users via audio channels. The system was designed for up to four users, again working on the DiamondTouch tabletop device. An experimental task involving the construction of a soundtrack to a movie using 17 representative clips from the movie and 34 icons representing songs chosen from a music collection, was given to 16 recruited participants, divided into four groups of four.

There were two conditions within which this task was completed – the "private sound" condition, where captions and songs were played over individual earbuds and the "public sound" condition, where captions and songs were played over a single shared speaker. In general, it was found that users preferred the "private sound" condition due to a number of factors. Users found that it was easier to complete the task in the private audio condition and it reduced table dominance by group members. Surprisingly, having a private audio channel did not reduce user communication. These factors led to better quality results and greater efficiency in task completion.

2.1.3 Human-Computer Interaction Summary

In this section, we have given an overview of principles and theories in single-user HCI, such as various usability tests (Shneiderman and Plaisant, 2005) and design principles (Norman, 1998; Nielsen, 1993) that can be used in the design of systems and their respective interfaces, which meet the needs and expectations of their target audience. These principles and guidelinesprovide the groundwork for research in multi-user HCI and we discuss how these principles and guidelines are incorporated in our system interface designs in Chapter 4 (Section 4.3). We have seen that multi-user HCI brings a complexity that has not been previously experienced in a single-user context, since we do not solely take one user's interaction with a single PC into consideration, but rather multiple people working with each other and with a computing system. This complexity comprises issues regarding territoriality and orientation, awareness, widget placement, division of labour, coordination policy and different input and output modalities available. While we have seen that research (e.g. Tse et al., 2004, Ringel Morris et al., 2006) is ongoing with regards to these issues in multi-user systems, there is still a need to further and improve such research, as groupware technologies become more popular, prevalent, diverse and sophisticated. In the next section, we introduce a number of such groupware technologies, which can be grouped into four different categories. These are discussed, along with examples, below.

2.2 Groupware Technology

Groupware, also known as computer supported cooperative work (CSCW), refers to technology that supports the tasks of groups and teams, either for work or entertainment purposes.

"Computer-Supported Cooperative Work is concerned with the need also to support multiple people working together using computer systems" (Greif, 1988).

This area of computing has been receiving much attention in recent years, particularly with the advent of Single Display Groupware (SDG). Exciting new technologies are emerging that effectively support the needs and wants of groups of people, working collaboratively (or against each other in the case of competitive games). Such technologies have sparked a huge interest among academia, industry and the general public and annual conferences have been organised to describe and demonstrate advances in research made in this area e.g. CSCW, IEEE Tabletops. This research encompasses effective technologies, interfaces, applications and guidelines that best support group-work (Ringel Morris et al., 2004b). Many of the technologies developed have just recently become commercial products e.g. Microsoft Surface (Section 2.2.4) and we will soon see these deployed in restaurants, offices and homes.

Designing a groupware system differs greatly from designing a single-user application. As described earlier in Section 2.1, designers need to take into consideration factors such as the background, skills, culture, motivation for use of the application, workstyle differences and social position of the members of the groups the application is targeted at. An application that is only acceptable or beneficial to only one member of the group, will inevitably fail.

Grudin, 1994, identified eight challenges for developers of groupware applications. In this article, the author looked at both successful (e.g. TeamFocus, Lotus Notes) and failed (e.g. automatic meeting scheduling applications) groupware applications; supplied and suggested reasons for their respective successes or failures, and devised a list of methods that developers could follow, which would increase the probability of their application being successful. Understanding the differences that exist between groupware applications, single-user applications and organisational information systems was something he considered to be essential in designing groupware technology.

Groupware technologies have been, and are currently being, designed to support a multitude of group-work situations. Ellis et al., 1991, categorised collaborative interfaces and technologies into a time/location matrix consisting of four types. This categorisation is shown below in Figure 2.7.

	Different Place	**Same Place**
Different Time	Distributed Asynchronous	Collocated Asynchronous
Same Time	Distributed Synchronous	Collocated Synchronous

Figure 2.7: *Four categories for groupware interfaces*

This categorisation has also been discussed in depth in (Shneiderman and Plaisant, 2005). It is this categorisation that we will use to describe various groupware technologies in the remainder of this section. Each of these categories surround situations that users are in and hence the technology used must support these situations. We provide an extensive description of collocated, synchronous technologies, since it is this category of groupware that we have studied in this thesis. We selected this category of groupware technology to focus on as this is relatively new in comparison to the other three categories and many of the technologies available that support collocated, synchronous work are research prototypes. We hoped that our research would contribute to the design of such technologies and systems, by noting the aspects that users like and/or dislike.

We proceed with a description of various technologies that have been designed and developed to respond to the needs and issues presented by each category.

2.2.1 Distributed Asynchronous Groupware

Several distributed asynchronous (also referred to as *Different Place, Different Time*) technologies are described below, along with the benefits and challenges that each one presents.

(1) Bulletin Boards Systems or NewsGroups

These are network-based and are generally divided into general topics of interest, which are further divided into specific sub-topics e.g. a general topic might be entertainment and a specific subtopic of this might be movies. People who join a group can submit queries, responses and comments about such topics. A query or comment submitted by a user, which receives a response(s) is known as a *thread* i.e. the initial query/comment and all of the responses (if any).

Bulletin Boards/Newsgroups provide a forum for people to discuss topics, and to acquire and supply information. Examples of Newsgroups include USENET (Usenet.com, 2008) and Google Groups (GOOGLE, 2008b). They have the advantage of being easy to join and access (i.e. wherever there is an internet connection), and facilitate information sharing and debate.

(2) Electronic mail (e-mail)

E-mail came into use in the 1960's (predating the Internet) and since its inception, it has been one of the most successful networked computer systems (Anderson et al., 2001). It is widely used by people for work, academic and personal purposes and enables people to compose, send, save and receive messages. It is increasingly replacing regular postal mail (so called “snail mail”), as the general population becomes more and more computer literate and network connections improve.

Attachments enable forms, documents, images and audio files to be transmitted along with an e-mail, further improving the facilities that e-mail provides, which in turn increases its popularity. Graphics can also be included in the body of e-mail messages. E-mail can act as a meeting scheduler and has recently become available on mobile phones.

However, issues exist with regard to privacy of the information transmitted, since e-mails are not encrypted. There are also the issues of spamming (i.e. unsolicited e-mail sent to possibly millions of accounts), phishing (illegal acquisition of sensitive information such as passwords) and virus transmission, which e-mail users must be extremely careful of.

(3) Text/Multimedia messaging

Text messaging is a service that enables people to send either private text or multimedia messages such as images from their mobile phone to another (GSM, 1985; Henry-Labordere and Jonack, 2004). It is often used as a means of communicating a person's location, or suggesting a meeting. It is also used to enable people to carry out relatively short conversations about general current occurrences or situations in their lives. Abbreviated words, acronyms and text phrases (e.g. “LOL”, meaning “Laugh Out Loud”, or the number four used in place of the word “for”) are frequently used.

(4) Cooperative schedulers and Calendars

Cooperative schedulers are useful for identifying the most suitable times to hold meetings (Ehrlich, 1987). Participants are requested to fill in the times that they are free to meet into the scheduler/calendar and the time-slot that is acceptable to most, if not all of the participants is selected. Examples of online scheduling software are Meet-O-Matic (Meet-O-Matic, 2008) and Google Calendar (GOOGLE, 2008a). IBM's Lotus Notes (Lotus, 2008) also provides facilities for group scheduling and calendars. They have a great advantage over traditional approaches in that the meeting requester can see at a glance when most people are available, rather than the time-consuming effort of contacting each person individually and attempting to respond to and cater for everyone's preferences.

Scheduling is not a trivial task. If there is no suitable time for all participants to attend, then issues such as status of those who can/cannot attend come into play. If there are time-slots where only one or two people that do not have a high status cannot attend, then these will more than likely be selected, possibly to the dismay of the non-attendees. In some cases, this can hurt morale and motivation. Therefore meeting organisers must be careful when selecting meeting times and take the opinions and requests of people who cannot attend into consideration.

(5) Central Information-Sharing systems

In organisations, central information-sharing systems enable people to submit documents and forms to a central repository that others in the organisation can conveniently access and use e.g. IBM's Lotus Notes (Lotus, 2008). This facility significantly increases the efficiency with which information can be obtained and used in everyday work or education scenarios.

2.2.2 Collocated Asynchronous Groupware

Same Place, Different Time groupware enables groups of people to work at a time convenient to them. In the examples below, multiple displays are used and group members can be scheduled to work on these at different times.

(1) i-Land (IPSI's Roomware©)

Roomware© integrates the real and virtual world (Streitz et al., 1999). With this technology, the authors viewed the world around us as the interface to information sharing and designed this technology based on that idea. See Section 2.2.4 below for more detailed information on Roomware©.

(2) CommChair

CommChairs (Müller-Tomfeld and Reischl, 1998) are movable chairs with embedded computers containing wireless network connections and individual stand-alone power supplies. They have the added appeal of providing a level of comfort to users equivalent to that of armchairs. There are two versions of the CommChair (Müller-Tomfeld and Reischl, 1998), one that contains a docking facility in the swing-up desk that is part of the armrest (this enables people to work on their laptop computers) and the other with a pen-based computer located in the swing-up desk. These computerised chairs implement the BEACH software (see Section 2.2.4) to enable people sitting in CommChairs to communicate with people in other CommChairs. The software also allows people sitting in CommChairs to remotely communicate and share information with people working on DynaWalls (see Section 2.2.4) and InteracTables ((see Section 2.2.4). However, group members need not act simultaneously.

(3) CAMM

CAMM or *Context Aware Mobile Messaging System* (Freye et al., 2007, p. 1) is "a context sensitive mobile messaging system which derives context in the form of physical locations through location sensing and the co-location of people through Bluetooth familiarity".

The CAMM system transmits four types of SMS messages or public digital signs to users' mobile phones based on their environmental contexts. These four types of messages are reminder messages, which are set to alert the user based on location and co-location of others; notification messages, which enable information to be transmitted as public or private messages; presence messages, which inform CAMM users as to the location or co-location of another person and a group storage facility, where information can be stored for future retrieval based on the users context.

To transmit these messages, CAMM requires location/co-location information as well as expiry times. It is Java based and consists of ubisense location sensors, Bluetooth co-location sensors, a Python series60 Symbian OS mobile phone application and a web interface. An Oracle database is maintained by the CAMM's server, and this server also makes use of an SMS gateway to transmit messages and communicates with the python application via XML data transmitted over GPRS.

2.2.3 Distributed Synchronous Groupware

Also known as *Same Time, Different Place*, this category of groupware enables people, located in geographically different places, to work cooperatively or converse, simultaneously. Daily tasks can be carried out in a much more effective and efficient manner as communication is improved among geographically distributed people. The following sections describe some of the technologies that support this type of group-work.

(1) Audio and Video Conferencing Systems

Audio and video conferencing systems provide a means of conducting both formal and informal meetings, where the participants are in different geographic locations (Bly et al., 1993). These systems are useful where travel to attend a meeting is undesirable or infeasible. These systems can be used to make arrangements with family and friends, to keep in touch with loved ones and acquaintances and also for more formal meetings, where specific topics must be discussed and negotiated.

Audio conferencing is simple to implement and effective at achieving its purpose. Participants to the conference can just add people to their telephone conversations and they can have discussions amongst each other.

Videoconferencing requires video cameras, a good network connection and computer software. They are the next best thing to face-to-face meetings. Specific times are scheduled among participants for Videoconferences and these meetings are conducted in rooms specifically set-up for such a meeting. Videoconferencing enables participants to visualise each others' reactions, body language and facial expressions, thus increasing awareness and reducing ambiguity in participant responses.

Desktop Videoconferencing (DTVC) is a relatively new phenomenon, which has provided users with a less expensive method of videoconferencing. At a basic level, all that is needed is the DTVC software e.g. Microsoft's NetMeeting and a regular video camera. This type of system can be used for both personal (to provide a more co-present feeling when communicating rather than just voice communication over a phone line) or formal purposes (such as in formal negotiations between professionals).

(2) Group Editor Systems

These can actually be either synchronous or asynchronous, but the most common tools are synchronous. They enable users to modify a file stored on a single computer simultaneously. Examples include the browser-based SynchroEdit (SynchroEdit, 2008, allows rich text documents to be edited), DocSynch (DocSynch, 2008, allows text files to be edited remotely and simultaneously), ACE collaborative text editor (ACE, 2008), Groove (Groove, 2008) and TatukGIS Editor (TatukGIS, 2008) (which enables GIS map files to be edited).

(3) Interactive Network Games

Traditional games consoles such as the Sony Playstation (Sony, 2008), Nintendo (Nintendo, 2008b) etc have now added support for playing games against remote opponents over the Internet (Borella, 2000). This networked gaming environment is described in more detail in Section 2.2.4 below.

(4) Online Chat/Instant Messaging Systems/Virtual Worlds

Online Chat enables users to socialise on the internet. Users can be represented by avators or icons and can assume different personalities in interacting with others. This can prove to be a serious issue, especially when children use such chat-rooms. Awareness about these issues is making people, particularly parents, more vigilant about monitoring their children's use of chat rooms.

Online chat can also be used to provide technical support services e.g. Dell offers real-time chat services with technical support staff to customers experiencing problems. These chats are recorded and subsequently e-mailed to the respective customer to enable them to look over and save what was said.

Instant Messaging (IM) systems are provided both internally within organisations (Herbsleb et al., 2002), (Nardi et al., 2000) and externally on the

Internet e.g. MSN (Microsoft, 2008a), Google Talk (GTalk, 2008). These networked systems enable real-time communication among users of these systems and provide a means for regular conversation between friends and loved-ones, as well as the real-time flow of ideas, questions and solutions to problems.

The most advanced technologies in this area are the so-called virtual worlds (Castronova, 2001), such as Second Life (Linden, 2008). There, people are represented as avators and can move in a virtual world, created by participants, and can interact with other online participants. There is a notion of 'place' within this world, which mirrors the notion of place in the real world. Online participants can interact with each other via text chat or via voice chat.

(5) ClearBoard

An early example of distributed networked surfaces is ClearBoard. In (Ishii et al., 1992), the authors described the architecture of 2 prototypes of ClearBoard (i.e. ClearBoard-1 and ClearBoard-2), a technology that enabled two people in two different locations to work on a surface while maintaining natural eye-gaze. ClearBoard-2 was a refinement of ClearBoard-1, based on user-testing and evaluation. The general architecture was similar for both versions, that being a display tilted at an angle of 35 degrees at both terminals, a video projector and a video camera. A CRT-based rear projection display, with a transparent digitizer sheet 80 cm x 60 cm in size was used. This was mounted to the surface of the flat-panel display. Users could interact with the screen using digitised pens.

The drawing marks made by the two people were captured by the video camera, positioned above the display. In addition, a reflection of the user was captured as a continuous video image. This was then transferred over a network and projected onto the other user's display. Users were hence able to "look" at the other person and see their expression as they were drawing, without having to significantly change their focus of attention. This was one of the goals of the system. It appeared to each user that the other was just behind the display.

The drawing system that authors tested this technology on was called TeamPaint, an application that ran over AppleTalk on Mac computers. Results from an initial study showed that the technology supported gaze awareness, that users felt that they were co-present, interacting with each other in a shared environment. The only significant problem encountered was the poor sensitivity of the digitised pen.

(6) Virtual Classrooms

There are many instances of online lecture facilities provided to people who take courses remotely. In (Baecker et al., 2003), the authors studied a webcasted virtual classroom, enabling students to view and hear the lecturer and interact accordingly. This technology also provided "video, audio, slide, and screen broadcasting; slide review; question submission; the automated creation of structured, navigable, searchable event archives, and automated data collection for evaluation". Such systems provide a more co-present feeling to remote students and enable them to socialise with their classmates.

2.2.4 Collocated Synchronous Groupware

As previously stated, we provide a more extensive analysis of this category (*Same Time, Same Place*) of groupware technology, as this is the category we have studied in this thesis. Firstly, we present a brief overview of non computer-supported collocated, synchronous group-work. Such environments have provoked ideas among researchers to extend these non computer-supported metaphors to computer-supported.

Traditionally, face-to-face meetings and group-work have been supported by real-world physical objects such as whiteboards, overhead projectors, acetates, tables, paper, pencils and physical objects of interest to the participants. Tables are an excellent and natural means of facilitating such group situations as they enable participants to retain eye contact, observe each other's body language and facial expressions, as well as enabling speakers to physically show what they are talking about to the rest of the group. Overhead projectors and acetates enable group members to present their work, sparking questions and discussion among the rest of the participants. Whiteboards support the flow and consideration of ideas e.g. brainstorming, as well as providing a means for people to explain, often by means of a diagram or table, what they are speaking about.

Meetings of this kind still frequently occur today. However, CSCW researchers have identified huge potential to design and build technology that supports group-work, such as that described above. Not only this, but planning activities, particularly spatial planning, would greatly benefit from such technologies, as locations being discussed could be viewed in real-time. Many researchers have exploited the table metaphor described above and designed and built interactive tables. Many of these are described below.

Colocated synchronous groupware can be broadly divided into single display groupware and multi-display groupware. We now give an overview of 13 different single-display groupware systems, followed by four examples of multi-display systems.

(1) Microsoft Surface

Microsoft Surface (Microsoft, 2007) is a multi-user, touch-sensitive tabletop device. Microsoft began research and development into this type of touch-sensitive technology in 2001. The Surface tabletop itself was announced to the general public in May 2007 with a release date set for November 2007.

Surface has a 30-inch scratch-proof acrylic display on the surface of a table-like form factor (see Figure 2.8). It can detect multiple objects and user touches and gestures, and it supports multiple users working simultaneously. Behind the display lies five infrared cameras that detect any objects/human touches on the display. These cameras then send this information to the operating system for processing. Surface runs on the Windows Vista operating system and also supports Wi-Fi transfer, Bluetooth, and Ethernet connectivity.

Figure 2.8: *Microsoft Surface (View, 2008)*

Surface features four key attributes:

1. Direct interaction: Users can interact with the tabletop naturally using touch and gestures, without the need for using a mouse or keyboard.

2. Multi-touch: It can handle lots of touch-points simultaneously from objects as well as users e.g. users can draw with more than one finger simultaneously.

3. Multi-user: Surface can handle multiple users and provides an intuitive means for mediating group face-to-face meetings and collaborative work.

4. Object recognition: Users can place objects on the table to trigger various digital responses. It can also be used for digital content transfer.

Surface is not widely available yet and consequently there are few groupware applications which it supports. However, this is likely to change with its launch as a commercial product. While Surface can handle multiple user inputs, it cannot detect which user has touched the screen.

(2) Lumisight Table

Lumisight Table is an interactive, horizontal, multi-user device (Kakehi et al., 2006). It has two properties, which differentiates it from many other interactive tabletop systems. Firstly, it displays four different views oriented to each of four users. Secondly, users' gestures and object placements are captured by a camera underneath the display.

The tabletop's display is made from the material Lumisty, which has opaque or transparent properties when viewed from angles within certain ranges (i.e. -25 to -55 and 25 to 55 degrees for opaque and -25 to 25 degrees for transparent). This property prevents certain users from seeing private or simply specific information, that other users can see - in other words, it filters out particular objects from certain users' views, thereby supporting the notion of private territories. Shared information can also be displayed. Underneath the display are a camera and numerous projectors, allowing these multiple images to be projected and displayed simultaneously.

Lumisight uses the opaque direction as a back-projection screen, thus allowing each user to see only the image from the projector in front of them. Layering sheets of Lumisty enables different images to be displayed in different directions. Two Lumisty films placed orthogonally to each other are used in the case of four users. A Fresnel lens, installed just beneath the Lumisty films, improves the quality of displayed images.

Due to the fact that the Lumisty material is transparent in a vertical path, images of objects on the tabletop's surface can be taken by the camera underneath the display, as the projectors display images. The locations and shapes of these objects and images can be subsequently translated to execute particular functions.

(3) Smartskin

Smartskin (Rekimoto, 2002) is a touch-sensitive and gesture-sensitive device. Its architecture is based on capacitive sensing. To realise this architecture, a mesh of transmitter (vertical copper wires) and receiver (horizontal copper wires) electrodes are positioned on the surface. The thickness of this surface can vary and can form a variety of shapes, depending on the purpose of the interface. The accuracy of establishing hand-shapes and the position of fingers is determined by the density of this mesh of wires i.e. the more dense the mesh, the better the accuracy. The system does not suffer from occlusion and lighting condition problems, which are frequently encountered in other interactive tabletop systems.

Two versions of the system were made, one a tabletop interface, the other a tablet-type interface. The tabletop interface was 80 cm x 90 cm in size and enabled multiple-hand operations and supported multiple users. It supported two interaction methods: mouse emulation with distance measure and shape-based manipulation.

The tablet (called *The Gesture Recognition Pad*) was 32 cm x 34 cm in size and used a finer grid pitch than the tabletop device, thus allowing it to detect hand contact more accurately. It employed three interaction techniques: multiple finger tracking; hand or finger shape as input and finally, identifying and tracking other physical objects

(4) The Multi-Touch Screen

The Multi-Touch Screen (Han, 2005) is a rear-projected, touch-sensitive device, which depends on *frustrated total internal reflection* (FTIR). It can detect multiple simultaneous touches and gathers true touch image information at high temporal and spatial resolutions. It is relatively inexpensive and can be scaled for larger installations. The screen is composed of a thick sheet of acrylic, that is 16 inches x 12 inches in size. The edges of this are polished clear as an optical waveguide.

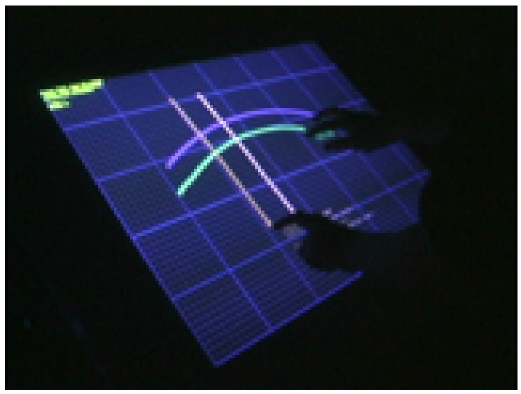


Figure 2.9: *The Multi-Touch Screen (Hans, 2008)*

High-power infrared LEDs are positioned directly against these clear edges in order to maximise coupling with total internal reflection. In addition to this, a digital video camera fitted out with a matching band-pass filter is placed orthogonally. The light remains within the acrylic and passes along predictable paths (total internal reflection) (Vaughan Nichols, 2007). When a user touches the surface, the light disseminates at the point of contact, causing the reflected paths to change. This process is captured by a camera placed below the surface. This camera in turn sends the information to image-processing software, which translates these touch points into their respective commands.

There are no ambiguity or occlusion issues with this technology. However, it is not as robust as technologies such as the DiamondTouch or MS Surface i.e. it is not scratch-proof. If scratched, the signal created when a touch is being registered is affected. This issue is currently being addressed by Han and new materials are being considered.

(5) Dialog Table

Dialog Table (Walczak et al., 2004) is a horizontal shared interface that enables people to interact with a museum's collection by using hand gestures. It was commissioned by the Walker Art Center by means of an international design competition and is permanently installed there. Its main objectives are to enable users to access the museum's collections and to educate users about art. Dialog Table promotes discussion among people on the movies, stories and 3D journeys from the museums archives that can be viewed on the table.

(6) DynaWall

DynaWall is a touch-sensitive, interactive computerised wall (Streitz et al., 1999). The display is 4.5m wide and 1.1m in height and covers one wall of a room entirely. The purpose of the DynaWall is to support the work of teams. Instead of placing sheets of paper on walls to organize information, information can be electronically organised and created collaboratively on a DynaWall. This functionality is realised through the BEACH software, created by (Tandler, 2004).

With this metaphor comes challenges in terms of object movement from one part of the display to another, particularly over long distances e.g. from one side of the wall to the other. To tackle this issue, Streitz and colleagues have implemented two solutions. Firstly, the user can select an object at one side of the wall, walk over to the other side of the wall and place the object at that location (without the need to be in contact with the wall the entire time). Secondly, the user can "shuffle" objects, which allows one team member to throw an object, to be caught by another team member at the other end of the wall.

(7) InteracTable

InteracTable (Streitz et al., 1999) is a mobile, computerised, interactive table that enables groups of between two and six people in size to create, display and annotate documents. It has a horizontal, touch-sensitive interface with dimensions of 65 cm x 85 cm and is a vertical rear-projection unit, 1.15 m high.

Users can interact with objects using a pen or their own fingers. Input can also be supported via a wireless keyboard. Objects are not oriented in any specific direction and such objects can be rotated or shuffled to enable users to view them at angles that are convenient to them.

(8) Touchtable^{TM}

This multi-user tabletop technology, created and supplied by Applied Minds Inc., is used specifically for spatial planning in Northrop Grumman Mission Systems (Northrop, 2008). It was designed to support collaboration among multi-disciplinary teams of people (i.e. analysts, planners, tactical staff), to enable them to view regions of interest, while simultaneously enabling access to data sources. The system dynamically changes projected images depending on where the users' hands are located and the associated movement of their hands. Specific gestures, once executed, invoke certain functions on this tabletop device, making the interaction more natural and intuitive e.g. moving two fingers away from each other zooms an image out. Different locations on the globe can be viewed at different degrees of granularity at any point in time. Information about each location can be displayed by touching the location point on the map once.

Two versions of TouchTable^{TM}were created – TT84 and TT45. TT84, as the name suggests, has an 84 inch diagonal surface, with an innate resolution of 1600 x 1200 pixels. TT45 has a 45 inch diagonal surface and has a resolution of 1920 x 1080 pixels. When used with TouchShare^{TM}software, multiple TouchTables can

Figure 2.10: *TouchTable^{TM}multi-user, interactive device (Northrop, 2008)*

be connected to each other to allow synchronous navigation by groups located in geographically different places.

(9) ROSIE Table

The ROSIE Table (New-Launches, 2007) is Savant's contribution to the genre of tabletop technologies and is the first to run on Apple Mac's OS X platform. It has a 40 inch touch-sensitive display at the surface of the table, which can be customised for media playback operations. It also provides home automation controls. Devices varying from multimedia streamers to networked security cameras can be controlled by the various applications developed for this tabletop.

The ROSIE Table's clever interface enables users to download and play multi-media material, such as obtaining photos from a digital camera, while also retaining complete control over the rest of the home applications. It is currently unavailable to purchase by the general public.

(10) DiamondTouch

The DiamondTouch is a multi-user, touch-sensitive, tabletop device developed at the Mitsubishi Electric Research Laboratories (MERL) located in Cambridge, Massachusetts (Dietz and Leigh, 2001). This tabletop allows up to four users to interact with it and with each other, simultaneously. Not only this, but it can uniquely identify each user when they touch the table's surface, as well as detecting multiple touches by a single user. The set-up is illustrated in Figure 2.11.

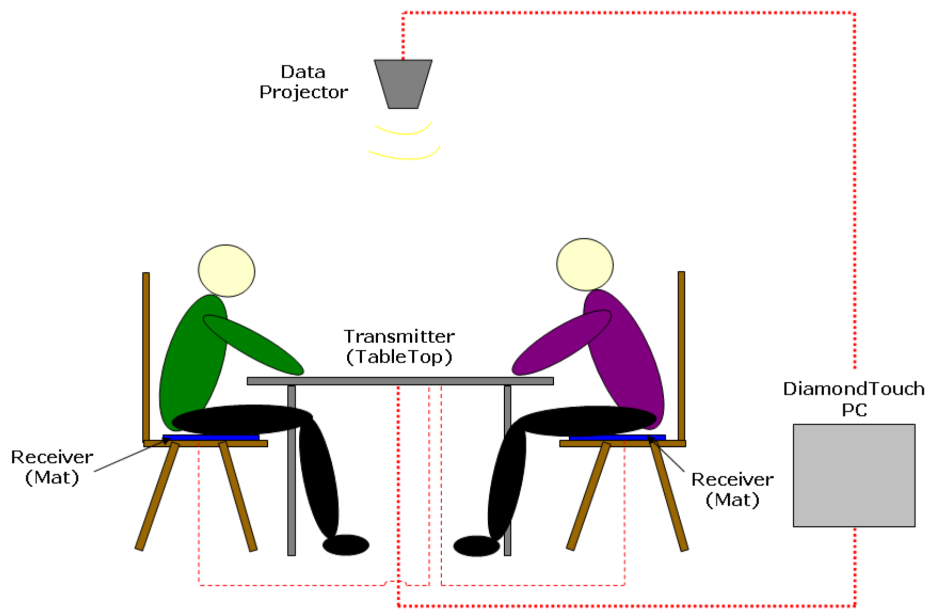


Figure 2.11: *DiamondTouch Setup*

The surface of the table consists of two layers of antenna, one layer arranged in rows and the other arranged in columns. These two layers are placed between row and column circuit boards with a thin insulator in-between. Each user sits on a mat containing a receiver, which is placed on a chair and connected to the underside of the tabletop. Once the user touches the table's surface, a capacitively coupled circuit is completed and a tiny signal runs from the transmitter in the table, through the user to the receiver in the mat and from here, back to the transmitter again. Any objects displayed on the tabletop's surface can then be manipulated by the user in a manner controlled by the software application. The Tabletop device is connected to a PC, from which the software applications are run. The interfaces to these applications are projected onto the tabletop's surface by a ceiling-mounted projector.

The DiamondTouch is extremely robust and can allow objects such as coffee cups to be placed on its surface, without interfering with operation of the table itself. The table also provides a natural interaction technique, that being the tabletop's recognition of a user's touch and gestures. This is one of the major benefits of the technology, as it is a more intuitive means of interaction – no intermediate devices such as mice or styli are required in order to interact with it. It is also reasonably inexpensive to produce and is durable as it does not require repair or re-calibration frequently.

Two versions of the DiamondTouch were built, one larger than the other. The DT88 has a 79 cm diagonal, while the DT107 has a 107 cm diagonal. Both these tables operate with an aspect ratio of 4:3. It is the larger table, the DT107, that we used in the experiments we carried out for this body of work.

Many applications have been built for the DiamondTouch using the DiamondSpin Software Development Toolkit (SDK) and the DTFlash toolkit. These handle multiple-user and multiple-gesture based applications and were designed and built specifically for the DiamondTouch. A mouse emulator was also built for the DiamondTouch, to ensure complete touch-based interaction with the tabletop.

The DiamondTouch is the system that we selected to conduct our research on. We had numerous reasons for choosing this technology over the other technologies listed above, which are:

1. The DiamondTouch can uniquely identify users working on the tabletop surface, hence it enabled us to track the users' interactions with the tabletop. This feature could be used to determine whether personality affected the interactions of users, whether those who interacted more had a better performance and whether territoriality was important in all of the systems we study.

2. The availability of a specialised software development toolkit (Shen et al., 2004), which made application development more convenient. This is described in greater detail in Chapter 3.

3. The robust nature of the tabletop made it feasible to carry out numerous experiments, without having to repair or replace parts of the hardware.

(11) MPX

MPX or Multi-Pointer X is a Groupware Windowing System (GWS) that allows numerous input devices to a PC to be used at the same time, e.g. keyboards and mice. It is really a modification of a Windows X Server to enable multiple users to interact with one or more applications on a PC simultaneously. It supports Single Display Groupware innately, together with a Multi-Pointer Window Manager. *"MPX and MPWM support an arbitrary number of true systems cursors, sophisticated floor control and per-window annotation overlay"* (Hutterer and Thomas, 2007, p. 1). The physical connection of the input device is invisible to both the application and the windowing system.

MPX can also be enabled on a Linux system. Natural and intuitive interaction is supported when a Linux MPX system is connected to a DiamondTouch.

(12) KidPad

This system (Stewart et al., 1998) was built on a previously reported MMM groupware study by Bier and Freeman (Bier et al., 1992). MMM groupware supported group-work by enabling multiple mice to be used as input devices to a regular desktop PC, as well as having individual mouse cursors and providing customised home areas, editors and menus for each participant. In (Stewart et al., 1998), the authors designed a collaborative version of a well-known drawing application called KidPix. Initially, 72 school children, aged between eight and twelve, were asked to work in pairs on the traditional KidPix system to draw a story cooperatively, without any software support for collaboration. In this case, each pair worked on a single PC with only one input device. The children were then asked to complete the same task using the KidPad system, which implemented an SDG architecture by supporting multiple input devices.

Results of this study demonstrated a large improvement in collaboration when using the KidPad system. The children had a more enjoyable experience, they experienced less collisions for the input devices, the amount of input was evened among passive and active participants and the children helped each other more and learned from each other. They even experimented by swapping input devices and using more than one input device each. The original KidPix showed that children frequently fought over the input device, they didn't listen to each other's advice, the child that didn't have control of the input device frequently diverted their attention elsewhere and expressed annoyance at not having equivalent control to the child who was using the input device.

(13) Video Game Consoles

Video game consoles (Berger, 2002) such as Playstation 3 (Sony, 2008), Nintendo Wii (Nintendo, 2008a) and Xbox 360 (Microsoft, 2008b) all feature single player and multi-player functionality for the games they support. These consoles generally have one set of hand-held controls/interaction tools for each player up to a maximum number (depending on the console), with all players looking at and playing on the same screen. Multiple players can either play simultaneously or be requested to take turns, but recently turn-taking has become less popular. Games consoles are generally used as home entertainment systems and with the widespread use of the Internet and the increased bandwidth available, games can now be played between individuals around the world. Since the majority of users still play in a collocated setting, we have categorised games consoles primarily as Collocated, Synchronous.

In addition to the above single-display groupware systems, there are four multi-display systems, that we will introduce here.

(1) IPSI's Roomware

Roomware (Streitz et al., 1999) refers to computer-augmented room constituents such as furniture, doors and walls, developed by IPSI in Germany. Integrated into this environment is information and communication technology, to move beyond the limitations of standard desktop environments. Roomware combines various computing technologies, such as Dynawall (Section 2.2.4), CommChair (Section 2.2.2), ConnecTable (Section 2.2.4), and InteracTable (Section 2.2.4), to create both a real and virtually augmented reality setting. Though the computer as a device becomes transparent, the functionality of the technology is freely and constantly available.

Some examples of roomware environments are the BEACH project (Tandler et al., 2001) and the iLand project (Streitz et al., 1999). Since each of the technologies used to constitute this computer-augmented room have their own displays e.g. ConnecTable, they collaboratively qualify as a multi-display environment.

(2) ConnecTables

These are networked devices that have been created to support the fluid transition between collaborative and individual work that occurs at various stages during face-to-face meetings (Tandler et al., 2001). These horizontal, pen-sensitive interactive devices were designed to tackle the "secondary interaction problem" (the technology's functionality) as opposed to the "primary interaction problem" (handling data input and outputtypes), as well as successfully integrating physical and virtual worlds. It is mobile and height-adjustable which enables users to assume a standing or seated position when working on it. The display can also be tilted to support different viewing angles.

Moving two ConnecTables close to each other creates a multi-user, shared workspace (i.e. from two separate workspaces). In this case, elements from one table can be accessed by a person at the other table – in other words, a homogenous display is created across the two tables. Objects can be exchanged between two people by moving them from one display to the other. To signify the creation of this shared multi-display workspace, the background colour of the workspace is changed. This type of connection is broken by moving the tables apart.

Documents can be reoriented to different people standing at different sides of the table. As well as this, several views of a document are supported and can be worked on in a specified view. The pen-based, interactive display has a 13-inch diagonal, with a 1024 x 768 resolution.

ConnecTables were made part of the BEACH ("Basic Environment for Active Collaboration with Hypermedia") roomware (Tandler et al., 2001).

(3) Office of the Future (HIT lab New Zealand)

Office of the Future (Billinghurst et al., 2005) uses a tabletop device to support the work of groups/teams of people. It enables people to place their laptops on the table and drag objects e.g. documents from the laptop screen to the tabletop display with a mouse.

Touch input is detected by a wide angled, low powered laser that shines across the display, with a camera mounted above it. When a user touches the surface, the laser shines a red spot over the person's finger and is tracked as it moves. This technique can be implemented to detect multiple fingers.

The tabletop display (public space), laptop (private space) and an additional wall display (presentation space) combine to give a very effective and satisfactory meeting environment, which was confirmed in an informal user study conducted by the authors.

(4) MERL's Multi-Device, Multi-User Environment

In (Forlines et al., 2006), the authors developed a software wrapper around the geospatial application Google Earth, to enable visualisation of geographical locations on multiple devices by groups of people. These devices consisted of three wall displays, a tabletop device (Dietz and Leigh, 2001) and a TabletPC. Different views of the information were achieved by running multiple instances of the Google Earth application on different machines in the same location, in this case four instances.

The main purpose of the wrapper was to ensure that all of the different views of Google Earth running were in-synch, with one instance of the wrapper running on each machine. The wrapper software also enabled multi-user input to the single-user Google Earth application, handled the rendering of some interface elements, enabled users to make annotations and share these annotations and also reconciled any conflicts that arose.

The main input screen here was the tabletop. Any changes made to the geospatial location being studied on the tabletop was transformed into different point-of-views (POVs) and sent over the network to the other instances of Google Earth running on the other machines. The POVs selected for each display were appropriate to the orientation of that display e.g. a bird's eye view for the tabletop display.

Small camera proxies on the table represented each of the wall displays and these could be modified to change the rotation and tilt of the image on these other displays. Wall displays could be tacked to "freeze" the image displayed, while other views were changed if the users so wished. Annotations made by users were geospatially registered by the wrapper application and subsequently displayed on each device.

The layers menu provided by Google Earth, which occupied almost 1/3 of screen real-estate, was removed from the public displays and instead put onto each individual's TabletPC, in order to prevent disruption to other users.

However, any changes made by an individual were still made visible on the displays. Also, changes made by two people synchronously to the location being examined were mediated by implementing a "driver's" policy, where the first one to touch had control. This control could be passed from one member of the group to another.

2.2.5 Groupware Technology Summary

In this "Groupware Technology" section, we have described the four dimensions of interaction that groupware technology must be designed for, namely Distributed Asynchronous, Distributed Synchronous, Collocated Asynchronous and Collocated Synchronous. For each of these, we have described several instances of computing systems, which support various kinds of groupware activities, from simple e-mail and bulletin boards to more complex multi-user, multi-display, collaborative environments. We gave a particularly lengthy description of technologies in the Collocated Synchronous section, as this was the category of groupware that we have chosen to focus on in our research. In particular, we provided a description of the DiamondTouch tabletop technology, which we chose to use for our user-studies. The reasons for this choice were also supplied, which were:

1. The DiamondTouch can uniquely identify users working on the tabletop surface, hence it enabled us to track the users' interactions with the tabletop. This feature could be used to determine whether personality affected the interactions of users, whether those who interacted more had a better performance and whether territoriality was important in all of the systems we study.

2. The availability of a specialised software development toolkit (Shen et al., 2004), which made application development more convenient. This is described in greater detail in Chapter 3.

3. The robust nature of the tabletop made it feasible to carry out numerous experiments, without having to repair or replace parts of the hardware.

In the next section, we discuss the area of Personality Psychology and its impact in the area of computing and the media. We discuss this area as Personality Psychology is an integral aspect of some of our main research questions, those being whether the personalities of our user-study participants have an impact on the performances, preferences and interactions when working in a team of two people (i.e. a dyad).

2.3 Personality Psychology

There are many schools of thought which attempt to explain or quantify differences between individuals e.g. (Allport, 1937; Freud and Riviere, 1927; Watson, 1929; Rogers, 1959; Neisser, 1967). These include the humanistic approach (Rogers, 1959), the behavioural (Watson, 1929) and cognitive approach (Neisser, 1967), approaches based on psychoanalysis (Freud and Riviere, 1927), and the dispositional approach (personality psychology) e.g. (Allport, 1937). We have chosen to focus on personality psychology for our work in relation to user preferences in multi-user, tabletop application design because we want to investigate whether Personality Psychology can add to our understanding of dyads human-human and human-computer interaction. Hence, this can give us clues about how to optimise the interface design of multi-user tabletop systems. We can also determine whether or not the combined personalities of dyads affect the performance, interaction and preferences of those dyads across simple competitive and collaborative games and tasks.

We now give a brief overview of some theories in Personality Psychology, by highlighting the contributions of some of the more influential theorists. While it is out of the scope of this thesis to describe the theories and models of these theorists in great detail, we provide this overview in order to familiarise ourselves with the various means of representing and quantifying the personalities of individuals (i.e. models of personality), so that we can pick the most appropriate and professionally accepted means of testing the personalities of our dyads. We follow this with some examples of personality profiling tests that can be used to measure people's personality type. Then we show how personality information can be used in the area of computing and the media, by presenting some examples of research in this area. Finally, we give a description of the effectiveness of groups in performing tasks when related to the group's personality composition, information that we will utilise later in our own studies.

2.3.1 Personality Theories

Personality psychology studies today can be categorised into two principal approaches represented by terms coined by the German philosopher Wilhelm Windelband: nomothetic and idiographic. Nomothetic approaches e.g. (Eysenck and Eysenck, 1964) or (Cattell, 1943) mainly focus on group data collection, with its principal aim being to predict group behaviour. In this approach, the impact of environmental and social issues are minimal and personality is seen as consistent, largely genetic and resilient to change. Idiographic approaches (e.g. Cooley, 1918; Mead et al., 1938; Rogers, 1959; and Erikson, 1969), on the other hand, focus on individual uniqueness and expansion of the idea of self. Personality development here is a procedure that is subject to variation, often shaped by the dynamics of interaction. In the past, psychoanalytic approaches (e.g. Sigmund Freud and neo-Freudians) and cognitive approaches (e.g. Kelly, 1963) took centre stage in personality psychology.

In the following subsections, we provide a short list of some of the major psychologists and various theories that have been influential in the development of personality psychology. While there are many more personality psychologists who have made contributions in this area, we picked these as they are widely known and have made the most major contributions to personality psychology as we know it today. These also have some impact on the work we report later.

Freud (1856 - 1939) was responsible for introducing the idea of the unconscious mind (Freud and Riviere, 1927). He believed that there were three major parts to the mind: the conscious, the preconscious and the unconscious (by far the largest part). These were in turn composed of subelements i.e. the ego, the superego and the id respectively.

One of his theories was that we are driven to satisfy physical and emotional needs - that this is in fact our Primary Process. The Secondary Process solves the problem of attempting to satisfy these needs based on our surroundings/environment.

Freud had a detailed theory[6] on neurotic anxiety and defense mechanisms that people had in response to anxiety. He used this theory to treat people with hysteria and other forms of psychological illnesses. According to him, reliving a traumatic event that patients had put out of their minds (i.e. bringing the event from the unconscious mind into the conscious mind), helped a person to cope with it, move on and become happier.

Carl Jung (1875-1961) was a Swiss psychologist who made great advances in personality psychology (Jung and Baynes, 1926) and was a founder of analytical psychology. He established the idea of personality traits as well as identifying personality components and archetypes. His early work was closely linked to that of Freud, though differences arose in their ideas of libido and religion, as well as the idea of psychological types.

Jung believed that the three basic components of personality were:

1. **The Ego** (the conscious mind).
2. **The Personal Unconscious** (this part of the mind is made conscious easily e.g. forgetting about something).
3. **The Collective Unconscious** (this is "the deposit of ancestral experience from untold millions of years, the echo of prehistoric world events to which each century adds an infinitesimally small amount of variation and differentiation." Jung, 1928, p. 162).

The first two components had been previously established, while he identified the third component (Jung and Adler, 1969). He also believed that there were four structural elements of the unconscious, also known as archetypes, which are:

1. **Persona:** This is the face individuals publicly use to respond to social situations.

[6]Here, a theory refers to " principles or procedures of any mode or field of cognitive activity, themselves considered as an object or branch of study" (OED-Online, 2008).

2. **The Anima and Animus:** These are sides to each person that represent Jung's view that human nature is effectively bisexual. The Anima is the feminine side of a man and the Animus is the masculine side of a woman.

3. **The Shadow:** This represents the unrealised opportunities in a person's life, both bad and good.

4. **The Self:** This element drives us to seek unity, harmony and wholeness among all parts of the personality.

Jung's theory on psychological types is his most renowned contribution to personality psychology. These types represent the manner in which people relate to the world around them and the way that they process information. Jung conceived the *Extravert* (interest in the outer world)/*Introvert* (subjective attitude) traits of personality, but soon realised that these were not adequate to measure personality completely. Hence, he added two other bi-polar traits, those being *Thinking/Feeling*, reflecting an individual's favoured mode of knowing the world and *Intuition/Sensing*, reflecting how a person perceives the world. A thinking individual defines objects or occurrences around her/him, whereas within a feeling individual, these objects or occurrences evoke emotional reactions. A sensing individual focuses on the existence of objects and occurrences, whereas an intuitive individual ponders on the cause and purpose of these occurrences.

The theory of personality traits, as conceived by Jung, provides the basis for the MBTI personality inventory (see Section 2.3.2 below).

The Big Five

More recently, personality psychologists have been converging to a consensus on a taxonomy of personality traits, as an accurate means of describing personality (Pervin and John, 1999).

This commenced when personality psychologists turned to natural language to build up a taxonomy of attributes. This involved the extraction of both descriptive and relevant terms from the dictionary. This process was guided by the lexical hypothesis, which conjectures that natural language has encoded most of the socially prevalent and pertinent personality characteristics (e.g. Allport, 1937). Following this method resulted in a finite set of attributes, although the size of this list was vast.

Allport and Odbert, 1936, identified four categories out of a list of almost 18,000 terms that could "distinguish the behaviour of one human being from that of another" (Allport and Odbert, 1936, p. 24). These categories were:

1. Traits
2. Temporary States
3. Highly Evaluative Judgements
4. Physical Characteristics

Later, (Norman, 1967), expanded these four categories into seven content categories:

1. Stable "biophysical" states
2. Temporary states
3. Activities
4. Social roles
5. Social effects
6. Evaluative terms
7. Anatomical and physical terms (as well as obscure and ambiguous terms that do not describe personality well)

Chaplin et al., 1988, applied prototype conception to traits, states and activities. Traits were seen as stable, long-lasting and internally caused, and needed to be monitored more frequently and across a broader variety of situations than states, before they were attributed to a person. States were seen as short-lived and externally caused.

Cattell worked on a subset of 4,500 words of the 18,000 terms of Allport (Cattell, 1943); (Cattell, 1945a); (Cattell, 1945b). He reduced this list further to a meagre 35 terms/variables, using semantic and empirical clustering, as well as his own reviews of available personological literature. He identified 12 personality factors in this list, which later became part of his 16PF questionnaire (Cattell et al., 1970). It was from this list that the *Big Five* was actually derived.

More simplified descriptions from 22 of Cattell's variables were built by (Fiske, 1949). The factor structures of these arose from self-ratings, ratings by peers and ratings by psychological staff members (which were almost identical). These factor structures resembled the future *Big Five*.

Tupes and Christal reanalysed correlations among eight different sample groups, from airmen with high school education to first year undergraduate students, in order to elucidate these factors. What they found was "five relatively strong and recurrent factors and nothing more of any consequence" (Tupes and Christal, 1961, p. 21).

Many personality psychologists replicated this structure from lists stemming from Cattell's 35 variables. These psychologists included (Norman, 1963; Borgatta, 1964; and Digman and Takemoto-Chock, 1981. The factors are summarised by Pervin and John, 1999) and are as follows:

1. Extraversion or Surgency (talkative, assertive, energetic).
2. Agreeableness (good-natured, cooperative, trustful).
3. Conscientiousness (orderly, responsible, dependable).
4. Emotional Stability vs Neuroticism (calm, not neurotic, not easily upset).
5. Intellect or Openness (intellectual, imaginative, independent-minded).

These became known as *The Big Five*, so-called to reflect the broadness of each factor (Goldberg, 1981). Goldberg chose adjectives for each factor that uniquely identified them.

These factors have high internal consistency and can be reproduced without difficulty. However, the fifth factor has the weakest replicability out of the five and if an alternative derivation is used to produce these factors, there is often a different label and description for this last factor. According to Costa and McCrae, the Five Factor Model "is the Christmas tree on which findings of stability, heritability, consensual validation, cross-cultural invariance and predictive utility are hung like ornaments" (Costa and McCrae, 1993, p. 302).

This Five Factor Model concedes four assumptions, that are generally implied in the area of personality trait research. These are:

1. **Knowability:** Personality is a subject that can be studied scientifically.
2. **Rationality:** People have the ability to comprehend themselves and others.
3. **Variability:** People are psychologically different.
4. **Proactivity:** The cause of human actions are sought within the person.

The Big Five In Other Languages

Similar to the manner in which the *The Big Five* was derived in the English language, personality factor models have also been found in Dutch, German, Italian, Hungarian, Filipino and Chinese as a result of cross-language research of the English derivation (De Raad et al., 1998). The most similar of these models to the English *Big Five* is the German version, which uses the same selection procedure as its English counterpart. Differences arise between these and the Dutch and Italian languages for the fifth factor which, as previously stated, has the weakest replicability. This is due to an alternative selection procedure used. The fifth factor in these languages represents Rebeliousness and Unconventionality, instead of Intellect and Openness. However, De Raad, Perugini and colleagues (1998) concluded that these seven languages support "the general contours of The Big Five model as the best working hypothesis of an omnipresent trait structure" (Pervin and John, 1999, p. 214).

Since this is the model of personality that is the most widely accepted and that the community of psychology professionals have converged on, this is the model of personality that we selected to use in our studies. For our research, we used this model to measure the personalities of our user-study participants. Measuring the personalities of our user-study participants in this way enabled us to answer many our main research regarding the effect that the combined personalities of dyads have on their performance, preferences and interactions.

The method for measuring the personalities of these individuals according to this model, is by means of completing a personality questionnaire, the answers to which were used to measure each trait in terms of a percentage (Johnson, 2008). This is described in greater detail in Section 2.1.2 below.

2.3.2 Personality Profiling

We now briefly describe three personality testing methods that have been widely used up to the present and we give a more extensive description of the Five Factor Model of personality which we have used in our study. We also highlight the importance of showing that personality profiling questionnaires used in research studies are valid and reliable.

Myers-Briggs Type Indicator (MBTI)

The MBTI is a widely used personality profiling tool, that categorises the personalities of individuals along four bipolar dimensions, giving 16 possible combinations. It was devised by Katherine Briggs and her daughter Isabel Myers-Briggs and is based on the theories of Carl Jung (see Section 2.3.1, also (Jung, 1971). The MBTI tool has been used by counselors in industry to boost relations between people at different levels in the chain of command of an organisation (e.g. Kennedy and Kennedy, 2004). It has also been used widely in the area of computing and engineering when studying the effect of personality in these areas (e.g. McDonald and Edwards, 2007; Da Cunha and Greathead, 2007). The dimensions measured by the MBTI inventory (Briggs Myers and Hauley McCauley, 1985) are:

1. **Extraversion/Introversion:** an orientation toward things outside oneself or a tendency to turn inward and explore one's feelings and experiences
2. **Sensing/Intuition:** whether a person is more prone to realism or imagination
3. **Thinking/Feeling:** whether a person is more logical and objective or more personal and subjective
4. **Judging/Perceiving:** one's orientation toward evaluating or perceiving things

However, recent research has questioned the conceptional foundations and psychometric properties of the MBTI (Gardner and Martinko, 1996). For this reason, we decided not to select this method for measuring the personalities of our user-study participants.

Eysenck's Psychoticism, Extravert, Neuroticism Model (PEN Model)

At about the same time as the *The Big Five* was emerging, Eysenck was devising his own model of personality (Eysenck, 1947). He believed that rather than just using factor analysis to develop a model of personality, other issues should also be taken into consideration, e.g. the fact that certain characteristics remain constant in an individual throughout their adult life. He believed that there were three factors that were central to an individual, that determined their behaviour and that any remaining factors could be derived from these three.

He initially derived a two-factor model, these factors being Extraversion (equivalent to Cattell's sociability and assertiveness and the Extraversion factor of *The Big Five*) and Neuroticism (equivalent to Cattell's emotional instability and apprehensiveness as well as *The Big Five*'s Neuroticism factor). Later he added a third - Psychoticism, defined as "a tendency toward psychopathology, involving impulsiveness and cruelty" (Freidman and Schustack, 2006, p. 295). The Psychoticism trait reflected Cattell's tough-mindedness and shrewdness factors and rated low on *The Big Five*'s Agreeableness and Conscientious scales.

Five Factor Model

Costa and McCrae began to derive their model of personality in 1976 (Costa and McCrae, 1976) with cluster analysis of Cattell's 16PF (see Section 2.3.1, Cattell et al., 1970). From this they started to develop the NEO Personality Inventory with just three of *The Big Five* traits (1980) - Neuroticism, Extraversion and Openness to Experience. The NEO Personality Inventory had scales to measure six facets for each of these three factors. However, in 1992 they published a revised NEO Personality Inventory (NEO PI-R), consisting of 240 items that measured six facets of all of *The Big Five* personality factors (Costa and McCrea, 1992b). These factors and their respective facets are illustrated in Table 2.2.

Several studies have been conducted that showed convergence between these scales and the *The Big Five*, though the idea of Openness appeared to be broader than the Intellect or Imagination factor, which was emerging from the lexical analyses at the time. These scales were shown to be internally consistent, temporally stable, and were convergent and discriminately valid against spouse and peer ratings. They could also be closely reproduced in a number of different languages.

In addition to this 240-item inventory, a shorter 60-item inventory was also developed by Costa and McCrae, based on an item-factor analysis of the longer

The Five Factor Model	
Extraversion versus Introversion:	Gregariousness (sociable) Assertiveness (forceful) Activity (energetic) Excitement-Seeking (adventurous) Positive Emotions (enthusiastic) Warmth (outgoing)
Agreeableness versus Antagonism:	Trust (forgiving) Straightforwardness (not demanding) Altruism (warm) Compliance (not stubborn) Modesty (not show-off) Tender-Mindedness (sympathetic)
Conscientiousness versus Lack of Direction:	Competence (efficient) Order (organised) Dutifulness (not careless) Achievement Striving (thorough) Self-Discipline (not lazy) Deliberation (not impulsive)
Neuroticism versus Emotional Stability:	Anxiety (tense) Angry Hostility (irritable) Depression (not contented) Self-Conscious (shy) Impulsiveness (moody) Vulnerability (not self-confident)
Openness versus Closedness to Experience:	Ideas (curious) Fantasy (imaginative) Aesthetics (artistic) Actions (wide interests) Feelings (excitable) Values (unconventional)

Table 2.2: *The Five Factors* (Costa and McCrea, 1992a, p. 49)

inventory. This abbreviated version was reported to be adequately reliable (Costa and McCrae, 1993).

Personality Test Validity

Selecting an appropriate test to profile peoples' personalities, that has been proven to be reliable and valid is important in the area of software engineering. This is mainly due to the fact that frequently, inappropriate and unreliable tests are chosen to profile individuals in this area, showing a lack of understanding of the underlying personality theories of these tests by those selecting and dispersing them (McDonald and Edwards, 2007). If an unreliable test is selected for profiling individuals in software engineering studies, then the results and subsequent conclusions drawn from these studies may be unreliable. Hence, it is important that a suitable and valid test be selected.

In (McDonald and Edwards, 2007), the authors highlighted the importance of test validity and reliability and made some suggestions for personality testing in the area of Software Engineering. Personality testing to predict job performance, or to identify good/bad software engineers (using tests such as MBTI) is not a good idea, as according to the authors, other factors such as work experience also have an impact. To identify which personality traits relate to software engineering abilities, a more suitable approach would be "to use a trait-based instrument (such as the 16PF), where comparisons to a normative sample can be made" (McDonald and Edwards, 2007, p. 2).

The authors also identified research studies and online tests that did not display data on their reliability, validity or correlation to the underlying theory they were using. For example, www.humanmetrics.com is a widely used web-site that profiles individuals based on the MBTI, but does not supply this fundamental statistical information and so cannot be said to be accurate by those who use it. To ensure validity, the authors here advise that researchers who plan on profiling the personalities of study participants become qualified testers or team up with a qualified tester. They should use publications that inform their users of the test itself, the administration process, who the professional testers are, how the feedback is to be imparted, and whether the types in the work are reported or proven types.

2.3.3 Previous Work on the Influence of Personality in Computing and in the Media

Some research has been conducted that examines the effects of peoples' personality types in the area of computing (e.g. Reeves and Nass, 1996; Karsvall, 2002; Brinkman and Fine, 2005; Rutherfoord, 2006; Da Cunha and Greathead, 2007). The studies listed, however, have utilised a single-user, single-PC condition to study such effects. Our study, in this thesis, appears to be the only one thus far that attempts to identify a relationship between users' combined personalities and their preferences and effectiveness in a collocated, multi-user computing task. In the remainder of this section, we provide a review of these previous studies, which demonstrate relationships between a user and their task-effectiveness and/or preferences with a single desktop PC, in different situations or in the case of different tasks.

Da Cunha and Greathead, 2007, conducted a study with 42 second year university undergraduate students to see if their personalities had any relationship to their effectiveness at completing a code review task. They used the MBTI personality test to profile the participants. They focused on the Sensing/Intuition and Thinking/Feeling dimensions as these were found to be the most important for relating to problem-solving capabilities and career choice. NT (Intuitive, Thinking) individuals were found to be better at the task of reviewing 262 lines of Java code (9.10) than the non-NT individuals (6.14) on average. SFs (Sensing, Feeling) achieved less than half the score of NT participants on average.

There are several reasons why NTs may have scored better than other combinations of these two bipolar dimensions. For instance, NTs are recognised as being best at solving problems within their area of specialised interest (Briggs Myers and Hauley McCauley, 1985). Another reason for the better results demonstrated by NTs is motivation. Since it is this type of task that NTs like, they may have been more motivated to do the task, in comparison to SFs who may not have been interested and hence, didn't apply themselves to the same extent.

Karsvall, 2002, studied the Graphical User Interface (GUI) preferences of individuals in relation to their personalities. He focused on analysing users' preferences with respect to the *Extraversion* trait and used Costa and McCrae's inventory to measure each participant's level of *Extraversion* (see Section 2.1.2, Costa and McCrae, 1993). Karsvall had three different variations of an iTV programming interface: a neutral design, an extravert design and an introvert design. The neutral design was simply the original interface as it was, without any modifications. The extravert interface used highly contrasting colours, more saturated hues and windows were given stronger lines and square shapes. The introvert interface used lower contrasting colours, de-saturated hues and weaker, more rounded frames – in general it was much more subtle. He hypothesised that participants would prefer the interface that matched their level of *Extraversion*.

The results of user tests conducted with these interfaces showed that the majority of participants (11/24) preferred the neutral interface; next came the extravert interface (8/24) and finally the introvert interface (5/24). Unfortunately, the skewed distribution of participants (only one was reported to be an introvert and 15 reported to be extraverts) failed to make these results statistically significant/reliable.

Brinkman and Fine, 2005, studied user preference of interface skins. Again, they tried to correlate each user's preference with their dominant personality traits. They used Costa and McCrae's Five Factor Model of personality to profile the participants (Costa and McCrea, 1992b, see also Section 2.1.2). Participants had to rate a series of Windows Media Player skins, as well as completing a Behavioural Inhibition Scale/Behavioural Activation Scale (BIS/BAS) and the personality inventory (IPIP-NEO). In the first study, the authors observed 99 peoples' skin selection preferences out of a set of 61 skins, randomly chosen from a set of 178 skins downloaded from the internet. The stages of initial and long term use were not studied.

The participants saw the skins in four stages: (i) the idle situation, (ii) the radio situation, (iii) the video situation, (iv) the CD/MP3 situation. The authors also obtained information with regards to the gender and age of participants. Participants were presented with the skins in sequences of 10 and rated these skins accordingly. The results showed that females preferred skins with "cute" images more than males and males preferred skins with themes more than females. The study confirmed the law of similarity attraction (which is a psychological theory that asserts that people like to interact with personalities similar to their own (Byrne and Nelson, 1965)) with regard to friendliness, cheerfulness, neuroticism, depression, self-consciousness, immoderation, excitement-seeking and humour, and their skin preferences. Also, their colour preference reported seemed to match user's personality in terms of their extraversion trait.

Due to the fact that only 17 participants returned the results of the IPIP-NEO test, the authors conducted a second study (Brinkman and Fine, 2005) using the results of this initial study. In this study, skins were not randomly selected from a downloaded set, but instead they were selected based on 17 categories of skin properties, such as their colour, shape, theme etc. Participants saw the skins in sequences of 17 and completed a reduced BIS/BAS questionnaire. The IPIP-NEO questionnaire was reduced to the 48 questions that related to the extraversion and neuroticism traits. Results showed high preferences for standard skins, but standard skins did not obtain the highest ratings supplied by participants.

Correlations in this study were reversed, with more gregarious participants being less likely to select colourful skins or skins with a small, friendly creature. Highly gregarious and highly assertive participants were also in favour of small skins. However, in general, the user preference dimensions reported and their correlations with their personality traits appeared to support the law of similarity attraction. In terms of other individual differences, younger users rated humourous skins more highly and again female participants preferred "cute" skins more than males.

In a later study (Fine and Brinkman, 2006), 40 users' interaction data with a web radio over a period of two months was gathered online, as opposed to a traditional lab setting with human observers. This was more natural, allowing users to use the application at a time that was convenient to them, as well as being more cost-effective. Users completed the IPIP-NEO questionnaire, a Short Test of Music Preferences (STOMP) and a more general descriptive questionnaire, which elicited information about the participant's age, background, gender etc. Correlations were found between personality traits (*Agreeableness* and *Neuroticism*) as well as behavioural measures, such as event numbers created in a session.

In (Agnihotri et al., 2005), a user study was carried out to determine if there was a relationship between users' personalities and their choice of summaries of broadcast television programming. The authors performed a factor analysis between the personality traits of the users against the features of their preferred summaries e.g. face presence, text presence etc. Personality traits were determined using the MBTI inventory, the Merrill Reed profile (Merrill-Reed, 2008), and the Brain.exe profile, completed by each of 59 participants. The results of the Brain.exe profile were later discarded since they did not reveal any statistical trends. The videos comprised of news, talk-shows and music videos. Each participant watched the video in its entirety and then watched a number of both audio-only, video-only and image-only summaries. They each selected one audio and one video summary along with four images that they thought summarised what they had just seen.

The results showed that only some personality traits were significant to specific genres and this significance varied across genres. For the news genre, the extravert/introvert traits from MBTI and the emote/control traits from Merrill Reed were significant, indicating that females and people with introvert and control traits generally preferred summaries containing actual reportage (e.g. text) rather than summaries containing anchorperson footage. In contrast, females and people with extravert and emotive traits preferred faces and disliked text.

In terms of talk-shows, intuitive people preferred the host speaking of the past and personal topics. Extravert thinkers favoured when guests on the video talked about their careers. People with control traits preferred text, bright portions and the verses of songs. The main difference that the authors found however, was that extraverts preferred experiencing direct content, rather than content mediated through an anchor or host. The opposite was true of introverts.

Personality in Computing and the Media

According to Reeves and Nass in their book "The Media Equation" (1996), the two most important types of media personalities are dominance/submissiveness and friendliness/unfriendliness. These relate to the first two factors of the "Five Factor Model" of personality - *Extraversion* and *Agreeableness*. Personality psychology and social psychology both apply to media studies. The authors inform us that media personalities are obvious and identifiable and that people like strong, reliable, identifiable media personalities, even though they may not like that personality once they have identified it. Inconsistencies in the personality traits and the appearance of a mediated character are disliked by most people. Media personalities influence our interactions and opinions of others. They also noted that, personality can be conveyed through simple things like an English sentence, by the language used i.e. dominant or submissive.

Numerous studies were reported in this book, which demonstrated personality in media. The most relevant study to the work in this thesis was the computer study (Reeves and Nass, 1996, Chapter 7, p. 89-99), in which the authors gave two computers a "dominant personality" and a "submissive personality". There were four differences in the computers which gave them these personalities. These four were:

1. The language style used for each computer i.e. assertions and commands versus questions and suggestions.
2. Confidence level in the comments expressed by the computer i.e. very confident versus not confident.
3. Interaction sequence i.e. dominant computer-initiated interaction versus submissive user-initiated interaction.
4. The names of the computers i.e. "Max" versus "Linus".

Forty-eight people who were experienced computer users were selected, half of whom had a dominant personality, the other half having a submissive personality. They used these computers to complete the *Desert Survival Problem*. This is a group problem-solving task, where group members are asked to imagine that a plane they've all been travelling on has crashed into a desert. There is no water visible, but twelve objects have been salvaged from the wreckage, which members must rank in order of their importance for survival (Lafferty and Eady, 1974). All participants completed the problem using a pen and paper. Then half of the submissive and half of the dominant participants completed the task using the dominant computer and the remaining participants completed the task on the submissive computer. They finished the task and completed questionnaires that evaluated the computer they used and the interaction.

The results showed that users perceived the computer with the dominant text as having a dominant personality, and the computer with the submissive text as having a submissive personality. Participants could say that each computer accurately reflected their own personality and preferred this computer. Participants believed their work was actually better on the computer that matched their personality. This proved the law of similarity attraction. However, participants denied any thoughts of a social relationship with the computer. The study showed that just simple cues could give machines a personality, and that advanced avatars, agents, guides, pictorial representations, or artificial intelligence were unnecessary to give technology a personality type. As the authors asserted, "Personality can creep in everywhere - the language in error messages, user prompts, methods for navigating options, and even the choices of type font and layout" (Reeves and Nass, 1996, p. 97).

A later study involved users completing the *Desert Survival Problem* on a computer with a consistent personality and then completing it on a computer that changed from a dominant personality to a submissive personality or vice versa (Reeves and Nass, 1996, Chapter 8, p.101-108). Results showed that users preferred the computer that adapted to their personality over time. They even perceived their work as being better on the computer that had adapted to their personality than the computer that exhibited the same personality as them throughout interaction.

A study that built on Reeves and Nass' work, sought to identify if people working with a tutorial to complete a task worked better with the tutorial that exhibited the same personality as them (Sayles and Novick, 2004). 35 students were selected to complete the *Adventure in the Amazon* problem, which had the same concept as the *Desert Survival Problem*. Their personalities were profiled using the Eysenck Personality Inventory (EPI) (Eysenck and Eysenck, 1964), which was administered at the same time as the actual task. People with a score of 0 were very introvert and people with a score of 20 were very extravert.

They completed the task on their own initially, and then with either the introvert tutorial or the extravert tutorial (this was decided prior to the testing). These tutorials simply used verbal cues to give them their extravert or introvert personalities. The results were analysed in three ways:

1. **Analysis 1:** People with personality scores above 10 on the EPI were considered extraverts, and people with scores below 10 were considered introverts. This meant that 13 extraverts used the extravert tutorial, 14 extraverts used the introvert tutorial, five introverts used the extravert tutorial and no introverts used the introvert tutorial.

2. **Analysis 2:** The median of all EPI scores was selected (15) and participants scoring below this were considered more-introvert. Similarly, participants scoring above this were considered more-extravert. This gave a more evenly spread sample, with four extraverts using the extravert tutorial, six extraverts using the introvert tutorial, 14 introverts used the extravert tutorial and eight introverts using the introvert tutorial.

3. **Analysis 3:** Linear regression was used to determine a relationship between a subject's level of improvement and the difference between the participant's personality score and the tutorial they used (0 for introvert and 20 for extravert).

The results of all of these analyses showed that none of the predictions made by the authors were statistically significant. No proof could be found to support a relationship between the success with which a person completed the task and the tutorial they used. Some explanations suggested were the small sample size; the fact that just one cue gave the tutorial its personality; the idea that people naturally improve when doing a task a second time, regardless of the personality of the tutorial used; or the possibility that the law of similarity attraction only applies to extraverts i.e. that introverts are attracted to extraverts rather than introverts.

In (Shim and Paul, 2007), the authors conducted a study involving 381 students to determine their level of attention to five genres of television programmes. They used Eysenck's three dimensional PEN model (see Section 2.1.2, Eysenck, 1947) to profile the participants. In the analysis of the data gathered, participants' personalities were divided in two - high and low, with the mean score serving as the cut-off point. From this division, the authors learned that the *Extraversion* trait was negatively correlated to the *Neuroticism* personality trait.

Participants rated the level of attention they maintained when watching Talk Shows, Soap Operas, News, Crime Dramas and Reality Programmes on a scale from 1 - 7, 1 being "not at all" and 7 being "very much". Data was gathered for a period of one month, from February 2007 to March 2007. Results from the analysis of this data showed that *Neuroticism* was the most significant trait related to media consumption, possibly stemming from emotional instability and lack of social adjustment associated with this trait. *Extraversion* was related to reality shows - high extraverts paid more attention than low extraverts.

Psychoticism was found to be negatively related to attention to news and reality shows, with low psychotics paying more attention to these two genres than high psychotics. In fact, high level psychotics were not found to be attracted to any of these television genres, probably due to the fact that television programmes generally contain less extreme content than movies. For instance, a previous study conducted by (Zillmann and Weaver, 1997) concluded that male participants with higher levels of *Psychoticism* were essentially more attracted to programming containing violence as a method of conflict resolution.

In the next section, we describe some research into the combined personalities of groups in work, educational and entertainment environments.

2.3.4 Group Personality Composition

We now discuss some previous research that has been conducted into group composition in terms of the personalities of the group's members. We draw information from these studies to provide us with a better understanding of what makes a good team, for use in our own work later.

In (Rutherfoord, 2006), a study was conducted with groups of people to determine whether groups with a heterogeneous personality composition were more productive and enjoyed working together more than groups with a homogenous personality composition. To test the personalities of the groups, Rutherfoord used the 126-question Keirsey Temperament Sorter (Keirsey, 2008), which categorises personality through following the Myers-Briggs scale (Briggs Myers and Hauley McCauley, 1985).

Twenty-two students participated in this case study, each completing the Keirsey Temperament Sorter, as well as answering questions concerning their prior work experience, ethnic background and the sex of each student. The class was broken up into six groups: three control groups (homogenous in personality composition) and six experimental groups (heterogenous in personality composition). Group sizes ranged from two people to five people and each group had at least one person with no experience with computers and at least one person with more than five years computer experience.

Their task was to develop a game management system for an Athletic Association. Results of this test showed that the homogenous control groups experienced more problems on a personal level, rather than experiencing technical problems. The heterogenous experimental groups conveyed a more broad and varied style of problem-solving, displayed much more interaction and discussion about alternative solutions, devised more creative and effective ideas and worked together outside class hours more. In general, heterogenous teams were deemed to be stronger and more effective than homogenous teams.

Gorla and Lam, 2004, distributed a questionnaire-based survey to 92 employees from 20 small software development teams (from three people to seven people in size), to determine what combinations of personalities resulted in the best performing teams. The survey elicited information about the amount, quality, effectiveness and efficiency of the work done, as well as the frequency that the schedule and budget were adhered to. The personalities of the participants of the study were also profiled using the Keirsey Temperament Sorter, which measures MBTI.

Results showed that team leaders with intuitive, feeling and judging traits performed better. Heterogeneity of personality between the team leader and the team members, particularly in the extravert/introvert and intuitive/sensing dimensions, proved to be more successful, though heterogeneity among team members had no significant effect. Thinking type systems analysts performed better, as their roles incorporated more tasks than in a larger team. Extraverted programmers performed better than introverted programmers. Diverse expertise and an appropriate means of sharing this information was also important.

In (Balthazard et al., 2004), the authors studied the performance of 63 virtual teams, composed of 248 MBA professionals, with respect to each members' level of expertise and extraversion, as well as the interaction style employed by the group. At the outset, the authors believed that the interaction style employed by a group mirrored the communication traits of each team member combined, these traits being rooted in each of their personalities.

Prior to this work, it was found that team performance generally depended on the expertise of the team members, with teams performing slightly worse than the most expert team member and better than the average team member (Hackman and Morris, 1975; Bottger, 1984; Cooke and Kernaghan, 1987). However, this was only the case if the group adopted an interaction style that facilitated the communication and sharing of this expertise. Teams that adopted constructive interaction styles produced better quality solutions and better solution acceptance than teams that had a passive/defensive or aggressive/defensive interaction style.

The quality of communication was also found to be one of the keys to successful team performance. Highly frequent communication, initiation of contact, positive tone and appropriate feedback style helped to establish trust (Jarvenpaa et al., 1988 and Iacono and Weisband, 1997). In this study, the authors postulated in this study that the group interaction style "is more a function of personality traits than knowledge" (Balthazard et al., 2004, p. 7).

Participants first completed a Five Factor Model profile and later an online "Ethical Decision Challenge" (Cooke, 1994), first individually to determine each person's level of expertise. They then completed it as a group that was randomly constructed. Communication between group members took place via an online chat and conference tool. When the task was complete, participants completed two questionnaires: a *Group Style Inventory*™, which assessed interaction behaviours within the group and a group process questionnaire, which assessed process satisfaction and "buy-in" into the consensus solution.

Analysis of the results showed that the best member's expertise was positively associated to the group performance, but negatively associated with team synergy. In other words, team performance was best predicted by expertise in the group. In general though, it was mostly the interaction style of the groups that had predictive power on the contextual outcomes in virtual environments and not individual personality or individual expertise. Variations in levels of extraversion in the group seemed to cause mainly negative characteristics. Extraversion was found to be an important trait to promote group interaction and teams with lower variance in extraversion levels did better.

A study was conducted in (Ludford et al., 2004) into online communities and the factors that promote participation in these groups. The online community, consisting of eight groups in the study, used a movie recommender system. Uniqueness and similarity combinations were tested to see which condition resulted in the largest participant contribution. Here, uniqueness was defined as a function of the discussion topic, the participant's movie ratings and the discussion groups movie ratings. Four of the groups were sent a weekly e-mail informing them of the unique perspective they could bring to the current discussion group. The remaining four groups acted as control groups. Groups were defined as "similar" if they typically watched the same movies and agreed on their reviews of these movies. Dissimilar groups either just watched different movies, or disagreed on movies the had watched. Again, there were four similar groups and four dissimilar groups.

Upon analysis of the results, it was found that dissimilar groups that were supplied with uniqueness information contributed more to these online communities. These results were contrary to the authors' hypothesised results, in that dissimilar groups participated more than similar groups and this diversity was significant. This is also contrary to many theories and studies completed in social psychology e.g. the "similarity attracts" phenomenon.

2.3.5 Personality Psychology Summary

We have seen from the above, that personality has a rich and complex effect on the workings of groups. We first briefly described some of the most renowned theories in the area of personality psychology. We followed this with a brief description of personality profiling tools, though we provided a more extensive description of the "Five Factor Model", which we used in our work. We referred to some previous research conducted with regards to personality in computing and the media. Finally, we discussed some research into group composition with respect to personality, which is highly related and significant to our work. In the next chapter, we introduce our work and some preliminary experiments we undertook prior to the main body of experiments.

2.4 Chapter Summary

In this chapter, we have introduced the three areas of Human-Computer Interaction, Groupware Technology and Personality Psychology, each of which forms the background to this body of work. We have described each of these fields of research, as well as the challenges that have been addressed and challenges that are currently being faced, in particular in the area of multi-user HCI. We commenced the chapter with an overview of HCI with a special emphasis on Multi-User HCI (Section 2.1.2), within which there are many issues that are currently being addressed, such as territoriality, orientation, awareness, widget placement, coordination policy and division of labour, particularly with regards to collocated synchronous technologies. However, more research into these issues on the design of multi-user interfaces is necessary, given the wide variety of multi-user technologies and the other respective categories that they fall into i.e. distributed asynchronous, distributed synchronous, collocated asynchronous.

We subsequently described a number of different types of technologies that fall into each of these categories in our section entitled "Groupware Technology" (Section 2.2). Here, we briefly described these technologies and where appropriate, the technologies that they comprise, their capabilities and how they are/can be used. Finally, we gave an overview and literary review of the area of Personality Psychology (Section 2.3)including a description of some of the most influential models in this area (Freud and Riviere, 1927; Eysenck, 1947; Jung and Adler, 1969; Costa and McCrae, 1993). We have also provided examples of related research studies that have been conducted in both the HCI and Personality Psychology sections. Their impact on and relevance to our work will become apparent throughout the remainder of this thesis.

Chapter 3

Designing Video Retrieval Interfaces for Dyads

In this chapter, we give a description of our own preliminary work in designing user interfaces for a two-person, video retrieval system. Firstly, we describe the importance of information retrieval systems in the modern world of Information Technology. Next, we give a general overview of an international video retrieval benchmarking activity known as TRECVid, which we developed some of our system interfaces for. We then describe the technologies used to build and host our system. Subsequently, we describe two different interfaces that we developed for a multi-user version of our video retrieval system. We discuss the results of user-studies that we performed on these interfaces, which gave us an insight into the collaborative workings of dyads and their interface preferences on such a video retrieval system. We also look at their performances with respect to their personality. This work and its respective outcomes led us to the hypotheses that we list at the end of the chapter and that we explore throughout the remainder of this thesis.

3.1 Introduction

The popularity of search systems has exploded in recent years. With the ever-increasing amount of information available on the Internet, be it in text, audio, image or video form, a need arises to be able to access this information efficiently and effectively to meet the information requirements of each individual. Activities that allow us to access this information include information searching and browsing (Marchionini and Shneiderman, 1988; Ellis, 1989; Kwasnik, 1992; Wilson, 1999; Baeza-Yates and Ribeiro-Neto, 2008). Browsing is the activity of following links created, either manually or automatically, between information items. Most of the main internet-based search systems support browsing. For example, each video in YouTube (http://www.youtube.com/) has links to "related content", popular social networking sites are all about links between individuals and other people (e.g. Facebook, 2008, Bebo, 2008), and of course HTML documents on the World Wide Web were designed to be accessed by browsing embedded links.

Information searching is an activity, which used to be the preserve of a small number of information professionals, but has now become a worldwide and open activity for everyone. Present-day information retrieval systems have become very sophisticated in their underlying techniques for satisfying the information needs of their users, though their interfaces and essential functionality remains limited. Since searching for information, be it using Google, Yahoo!, MSN etc. has become part of our daily routines, both for work and leisure purposes, it is important that search systems return information that is relevant and useful to their users.

Marchionini (Marchionini, 1995) devised an Information-Seeking problem stages-of-action model, which has nine stages as follows:

Stage 1: Recognise and accept an information problem

Stage 2: Define and understand the problem

Stage 3: Choose a search system

Stage 4: Formulate a query

Stage 5: Execute the search

Stage 6: Examine the results

Stage 7: Make relevance judgements

Stage 8: Extract information

Stage 9: Reflect/iterate/stop

This nine-stage model is broad and all-encompassing, and we use it as a basis for our own perspective on the information seeking problem. Accordingly designers and developers of search systems should understand each stage of this model and support it, to ensure its overall success and the ultimate satisfaction of the information seekers.

Though many other models of information seeking exist (e.g. Ellis, 1989; Kuhlthau, 1992; Wilson, 1999), we chose Marchionini's since its simple nature allowed collaboration to be incorporated easily. By adding the element of communication to this model, we can propose the following two additional stages:

Stage 1: Recognise and accept an information problem

Stage 2: Define and understand the problem

Stage 3: Consult about which search engine is most appropriate to select

Stage 4: Choose a search system

Stage 5: Consult about what query to enter

Stage 6: Formulate a query

Stage 7: Execute the search

Stage 8: Examine the results

Stage 9: Collaboratively make relevance judgements

Stage 10: Extract information

Stage 11: Reflect/iterate/stop

As stated earlier, the topic of this thesis is how the personalities of end-users and the interfaces they are using, impact on the effectiveness and preferences of certain tasks, specifically when carried out by dyads. Here, by task, we refer to the tasks which we used for our five systems – two of which were competitive in nature, where a person must score the greatest number of points over their opponent in a bubble-popping game and a card-matching game, and the remainder being collaborative tasks; two where a dyad must collaboratively find all matching sets of cards in a collaborative card-matching game (the first having a speed rule and the second having an accuracy rule) and three collaborative interactive search tasks where a dyad must collectively find relevant segments of video to a predefined information need (two of these tasks had a time-limit and one where finding a quota of relevant segments was required). Both the purpose and the level of complexity of these tasks are summarised in the Table 3.1. Since our user tasks do not easily fit into the standard task complexity definitions as described by e.g. Byström and Järvelin, 1995, we devised our own level of complexity, by determining the number of decisions required to complete a task. In doing this we thus developed a new classification of task definitions into easy, medium and hard. For instance, our search tasks are typically subsets of other tasks (i.e. satisfying an information need could be undertaken in order to proceed with a more general decision task (a "genuine decision task")) . In our case, where only one or two decisions are required, the task is deemed "easy" (e.g. pop-a-bubble requires one to burst a bubble, memory game requires a user to select a card). Where three or four decisions are required, the task is deemed "medium (for our search task, at least four decisions are required (collaboratively) select a search query type (i.e. image,

Task	Complexity	Purpose
Bubble-Popping	Easy	Simple task to introduce users to the system and help overcome any anxiety about using a new technology
Competitive Card Game	Easy	To see whether people preferred an interface matching their level of extraversion
Accuracy and Speed Collaborative Card Game	Easy	To see whether some people worked better under certain constraints and whether this matched their task constraint preferences
Collaborative Search (Quota-Based)	Medium/Hard	To see how long people take to find relevant video segments and whether this relates to their personality
Collaborative Search (Time-Based)	Medium/Hard	To see how people perform when executing a search task in terms of their recall. Also, we wished to see if people preferred completing this task on an interface that matched their level of extraversion and whether this affected their performance

Table 3.1: *Task Complexity and Purpose*

text or image and text); (collaboratively) select a search query object to enter into the search system; (collaboratively) determine which functions to use in order to determine relevance; and determine whether the shot is relevant or non-relevant). Finally, a task where more than four decisions are required is deemed “hard”.

We purposely chose tasks that were quite different in nature, in order to determine whether there were personality traits that impacted performance and interaction when using the DiamondTouch regardless of the types of tasks to be carried out. If this proved to be the case, then this would be a contribution to the area of Personality Psychology.

In order to make some initial progress with this research, and to get some experience of end-user testing, our first venture in developing an interface for a collaborative activity was to develop an interface for a collaborative video search task. Traditionally, our work has centred around building successful video information retrieval (search) systems. In particular, we implemented our first collaborative interface techniques as part of the TRECVid benchmarking conference series (Voorhees, 2004). This involved carrying out a user-experiment, where eight dyads performed a series of twelve search tasks on two interfaces to a two-person video information retrieval system, in order to analyse their performance and interaction on both interfaces. A brief description of this overall conference series is provided in Section 3.2.1 below. The work described in this thesis was partially formulated and partly tested on a multi-user version of our video search system, designed for the TRECVid benchmarking workshop, which we describe throughout the next section of this chapter.

3.2 Físchlár-DT

Search systems have traditionally been designed for searching for and browsing through documents consisting mainly of text, and for browsing the output of such searches, namely ranked lists of such documents (Van Rijsbergen, 1979). More recently, image, audio and video retrieval techniques have been developed and are continually becoming more sophisticated. However, much improvement is still required if we are to have true content-based multimedia retrieval. For example, in order to retrieve information from spoken audio files, speech recognition systems are needed to transcribe the words spoken on the audio track, which can subsequently be searched using text retrieval techniques. Image segmentation, colour detection, edge detection, object character recognition (OCR), object recognition and face recognition techniques have all been utilised for image and video retrieval, as well as shot boundary detection, story segmentation and text transcription of the audio track (closed-caption transcripts) for video. It is out of the scope of this thesis to describe each of these techniques in detail. However, detailed descriptions of these are available from (Smeaton, 2007) and (Smeaton et al., 2005).

Within our research group, we developed the web-based Físchlár search system as a generic video management and retrieval tool (Lee et al., 2006) that retrieves digital video shots from digital archives of TV news broadcasts. A version of Físchlár was also designed for other types of television programming such as sports, movies and sit-coms, as well as a version to provide search, browse and playback of educational content by means of a digital video search system for Nursing students in our University (Smeaton et al., 2004; Gurrin et al., 2004). Of most relevance to us in this body of work is the Físchlár News system (Lee et al., 2000). In this system, each 30-minute TV news broadcast is broken up into individual stories, which are subsequently broken up into shots. Each shot is represented by an image called a keyframe. Figure 3.1 shows the standard web interface developed to the Físchlár News video search system from a desktop machine or laptop.

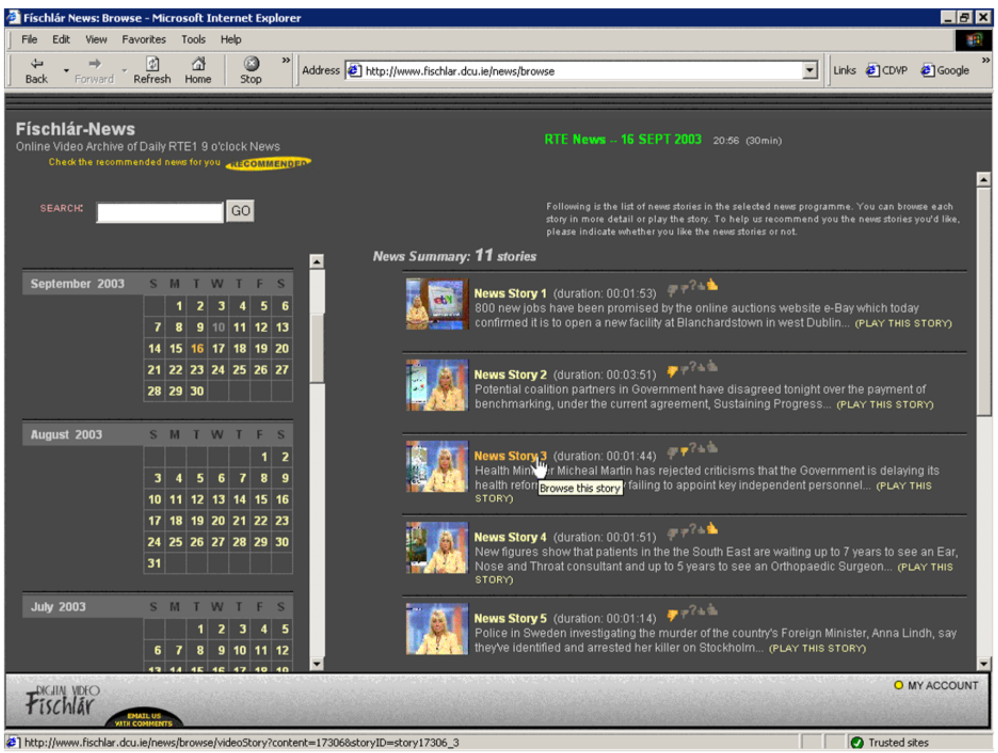


Figure 3.1: *Físchlár News search system*

As this figure illustrates, searching can be executed by typing a text query into the text box at the top left-hand corner of the page. Browsing is achieved simply by clicking on any of the dates in the calendar at the left-hand side of the page, each of which displays the respective news stories on the right-hand side. Each of these stories has a descriptive title, a brief text description of the story (the first two lines of the text transcript) and an image of the news anchor telling that particular story. It is also possible to rate the stories a user looks at in order to personalise that users' searches. This also allows the system to recommend stories it thinks may be of interest to them.

Some of these techniques have been evaluated and the system's performance compared against other video shot retrieval systems in the annual TRECVID workshop, which we now describe in more detail.

3.2.1 The TRECVid Benchmarking Workshop

TRECVID (Text REtrieval Conference for VIDeo, Smeaton et al., 2005) is a benchmarking conference series held annually, where participants compare video retrieval and analysis techniques and the performance of these techniques against each other on a shared test dataset. TRECVid was initially part of the bigger TREC (Voorhees, 2004) conference series, which focuses on text retrieval techniques. The entire TRECVid and TREC series of annual workshops and benchmarking tests are co-sponsored by the Defense Advanced Research Projects Agency (DARPA, 2008), which is a U.S. Department of Defense body, and the National Institute of Standards and Technology (NIST, 2008). In TRECVid, an archive of broadcast TV news is supplied to each participating group, as well as a list of topics that must be searched for, so that each group's system(s) can be fairly compared. TRECVid has a manual, automatic and interactive search task, in which interested groups can participate. Participants come together at an annual workshop held in Gaithersburg, Maryland, USA to present their systems and share their performance results. More information about the metrics used to measure performance in TRECVid is supplied in 3.2.3.

TRECVid follows the system-oriented approach to the evaluation of its search systems, including interactive search systems, which we focus on in this chapter (Smeaton et al., 2005). This type of evaluation centres on a well-defined, uniform task, of which there are many instances used to test a video retrieval system. Many instances are used to control for different levels of query difficulty. While this approach enables general inferences about the system to be made, it may result in focusing on just one quality aspect (Kraaij and Post, 2006).

In 2005, we decided to design a novel system for the TRECVid interactive search activity, that being a two-person video search system. Our reason for this was that we had become interested in the idea of collaborative search, an area that has been developed further in the group since then (Smeaton et al., 2008). Also, we had been contributing similar video search systems annually since the TRECVid conference series began and so we wanted to try a different approach to see if similar performance could be obtained by dyads completing the same task (Foley et al., 2005). All video search systems designed for TRECVid prior to this were single-user systems. We believed the idea of a two-person search system was interesting, as it presented new challenges that had not previously been addressed in this context. As stated in Chapter 2, we developed this system on the DiamondTouch tabletop system (Dietz and Leigh, 2001). We also used the DiamondSpin Software Development Toolkit, which we now describe.

The DiamondSpin SDK

This is a Java-based software platform and was designed and built by researchers at MERL to handle the arbitrary orientation and rotation of objects, such as documents, on rectangular, octagonal and circular tabletops (Shen et al., 2004). These are achieved through the implementation of a real-time polar to Cartesian transformation engine, where documents are displayed relative to one meaningful centre. Therefore, there is no specific position in which a document can be displayed. Documents on the table can be rotated, as well as the entire table surface.

The DiamondSpin toolkit has a well-defined API containing 30 Java classes and interfaces. It also uses pure JAVA 2D with JAI (Java Advanced Imaging) and JMF (Java Media Framework). This has the advantage of being platform-independent in nature, ensuring portability and extensibility for the toolkit, so the controlling PC can run Windows, Linux, etc. Traditional input devices are handled in DiamondSpin through the implementation of mouse input events. These input events are subclasses of Java mouse events, with the addition of fields, for example the unique user ID.

Within the DiamondSpin toolkit, a framework is provided for describing every component of the display in terms of both a polar orientation and a polar distance and constitutes two key concepts. These are:

1. Translation of the origin of a conventional Cartesian display (usually at the top left or bottom left corner of the TableTop) to the centre of the table.
2. 3 Degrees of Freedom (DOF) d, α and β for each element on the display.

Figure 3.2 (Smeaton et al., 2006b) below shows a document being displayed at a distance d to the centre of the TableTop O, at angle α. Each document can then be rotated around its own centre, or another point such as the point the user's finger is touching, using the β values.

Each DiamondSpin application uses four layers to display its composite elements. The lowest layer, layer 3 (Background Layer) contains non-interactive components such as background images. Layer 2 (One view Layer) contains elements which may potentially become active but are currently not receiving inputs. Layer 1 contains active components – components which are currently receiving input events such as being dragged, resized or rotated. Layer 0 (the Real-Time Polar to Cartesian Transformation Engine Multiple layer Layer) contains rotation sensitive components such as menu bars.

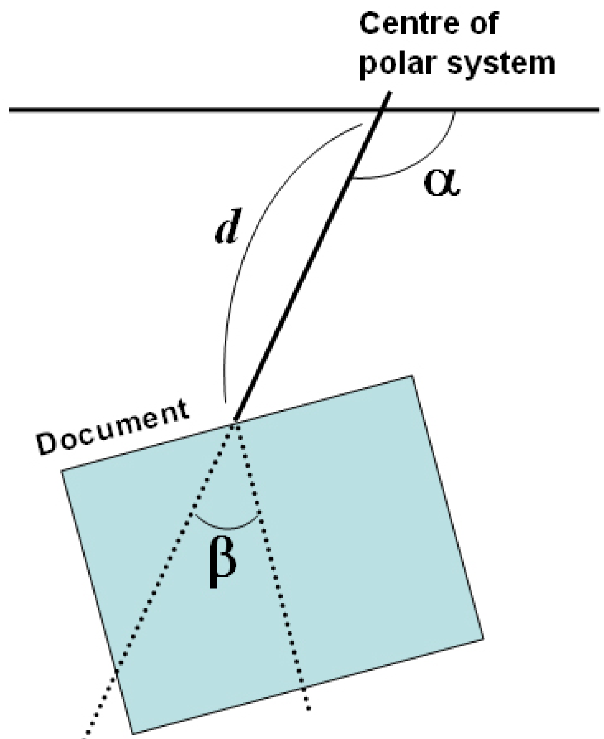


Figure 3.2: *Rotation and orientation of a document*

DiamondSpin uses a multi-layer representation thread with a multi-depth, multi-thread repaint methodology to handle multi-user concurrent interaction. Concurrent user input events are handled by multiple threads, one per user, with a unique thread name, which handle user interaction with tabletop elements. Another independent thread handles repainting the UI, displaying changes to the surface. The DiamondSpin API allows common Java factory components to operate on DiamondSpin's polar co-ordinate system, thereby allowing any UI component to be displayed at any orientation and rotated freely around the UI. The toolkit allows DiamondTouch application developers to use ready-made widget behaviours and common features for tabletop interaction, which enables them to concentrate on application specific interaction activities.

3.2.2 Description of The Físchlár-DT System(s)

In designing a two-person video search system for use on the DiamondTouch, it was necessary to consider issues with regard to multi-user HCI, as discussed in Chapter 2. These included awareness, territoriality, widget placement, orientation, division of labour and coordination policy. Taking these issues into consideration (Gutwin and Greenberg, 1998; Scott et al., 2003; Scott et al., 2004; Smeaton et al., 2006a), we decided to create two versions of our video search system for the TRECVid interactive task; the first focused on enabling the users to complete the task as efficiently as possible, while the second enhanced users' awareness of each others' actions. These systems were intuitively named "Efficiency" and "Awareness". We recall that the strengths and weaknesses of the TREC approach to the evaluation of information retrieval systems were discussed in Section 2.1.1, Chapter 2.

We decided not to implement a software-imposed coordination policy, but rather decided to allow users to use their own natural social protocols in order to coordinate their actions. This was so that user's true personalities would be apparent and allowed us to explore whether such policies applied in a technology-mediated context. Since video shot retrieval is a very flexible task, users were also free to decide how to execute the task themselves, rather than assigning roles.

We describe these two interface variations in more detail presently.

Físchlár-DT Efficiency Interface

Figure 3.3 shows the interface to the Efficiency variation of the Físchlár-DT system. Here, two people sit at each of the long sides of the rectangular table. One of the users can then type a search query into the text-box at the bottom right-hand corner of the screen. Typing is accomplished by double-tapping on the top of the search-box, which pops up a virtual keyboard, allowing the user to then enter text into the search-box. To send the query to the underlying search engine, the user then presses the "Search" button on the search-box. Out of the possibly tens of thousands of results, 20 keyframes representing 20 shots of TV news broadcasts are displayed around the table, the most highly-scored relevant ones being closer to the centre. Rotation, orientation and resizing of the keyframes is handled by the DiamondSpin toolkit (see Section 3.2.1 above).

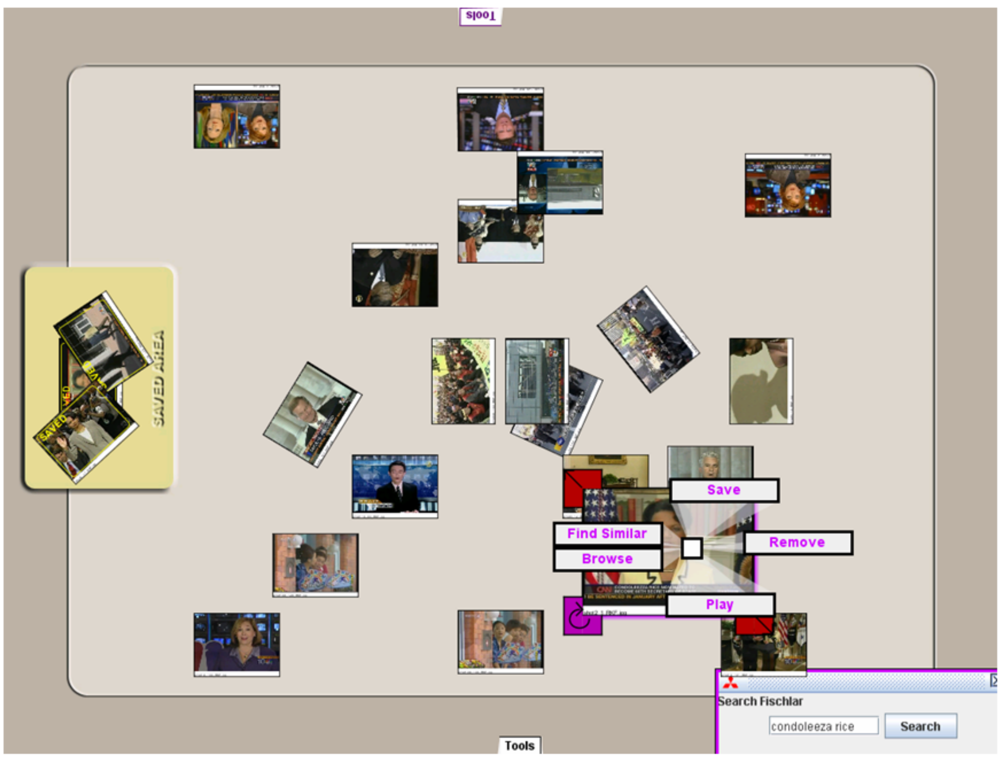


Figure 3.3: *Físchlár-DT - Efficiency version*

Additional functionality is available to a user by double tapping on a keyframe (similar to a double-clicking action on a mouse), which pops up a menu offering five functions: *Find Similar*, *Browse*, *Play*, *Remove* and *Save*. *Find Similar* executes a query to find similar keyframes to the one selected using a combination of the keyframe's colour characteristics and text from the spoken dialogue, if the text-box is populated. Further details of how this operates can be found in (TRECVid-Guidelines, 2005). *Browse* returns the next ten and previous ten keyframes in the news broadcast of the selected keyframe. *Play* enables the user to play that particular shot on an external monitor. *Remove* deletes the selected keyframe from the search for this search topic i.e. it cannot be returned again for this topic. Finally, *Save* moves the selected keyframe into the "Saved" area on the table, where it is stamped, deeming it relevant to the particular topic being searched for.

These five functions are standard features that have been used in TRECVid video search systems, although they are implemented in different ways by each participating site (e.g. Browne et al., 2003; Westerveld et al., 2003; Adcock et al., 2004; Browne et al., 2004; Heesch et al., 2004; Kender et al., 2004). These functions realise features of video search systems such as relevance feedback ("Find Similar"), context of the shot ("Browse"), text search, playback of the shot ("Play"), deletion of an irrelevant shot ("Remove") and the saving of a relevant shot for performance evaluation ("Saved").

By supplying each user with their own pop-up menu for each keyframe as it appears on the DiamondTouch, we believed that users could sift through the keyframe images at a faster rate as they could work silently and on a more individual basis (Fitts, 1994). However, a system such as this can also stifle collaboration as communication would be reduced and errors in coordination could rise, particularly in the case of the *Play* feature, where one user could play their video over their partner's. This is because actions could be executed discretely with a pop-up menu and so it is possible that both users could select the same option on each of their menus, thereby playing over each others' videos, overriding each others' searches etc. Our aim here was to see if this reduction in collaboration paid off in terms of the efficiency and accuracy with which the task could be completed. Our expectation here was that the efficiency version of the interface would enable users to perform better in terms of their precision and recall (i.e. they could sift through shots faster).

Físchlár-DT Awareness Interface

The Awareness system had the same functionality as the Efficiency system. However, rather than double-tapping a keyframe to pop up a menu of functions, five hot-spots were placed at different locations around the table, where users dragged their selected keyframes onto in order to execute the required functions. A sound is associated with each hot-spot after a keyframe has been placed on it, supplying feedback that the function has been invoked. This is shown in Figure 3.4.

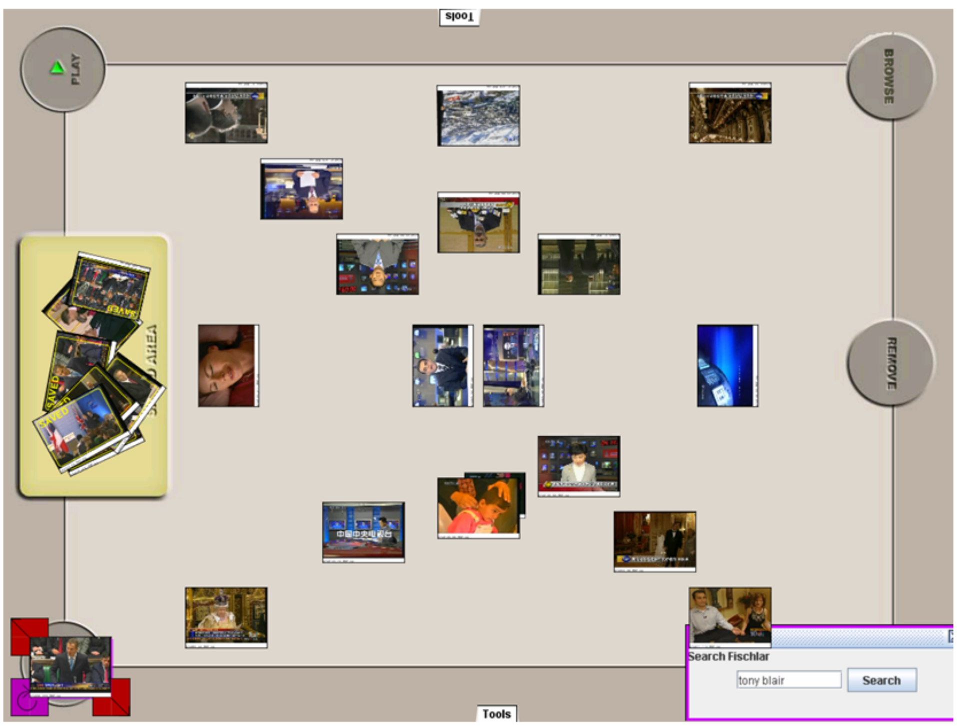


Figure 3.4: *Físchlár-DT - Awareness version*

By placing these hot-spots at different points around the table, and associating sounds with invocation of these functions, each user is made more aware of their partners' actions, due to the fact that they can see their partner dragging keyframes to these spots. Support for this can be found in (Gutwin and

Greenberg, 1998) and also in (Scott et al., 2003), where "collaborators are more likely to see another collaborator access an icon when using a touch-sensitive display since their whole arm is moving in space than when using a mouse when only a cursor moves" (p. 14). As a result, coordination would intuitively be improved, particularly with the *Play* function, as one user can see if the other user's keyframe is placed on that spot. Communication would be also increased as users must share the hot-spots and so cannot work as individually as in the Efficiency version. We verify this later in Section 3.2.3. Also, since some of the hot-spots are in each user's territory, social coordination rules can be implemented e.g. one user passing a keyframe to the other user to execute a function whose hot-spot is located in their territory (Smeaton et al., 2007).

The Search Task

To evaluate this system according to the TRECVid guidelines (2005), we conducted a set of user-tests with 16 users (8 pairs of users or dyads) (Smeaton et al., 2007). These individuals consisted of PhD student volunteers from within the CDVP research group and were paired based on those who worked closely together on related research areas. All of these participants worked in the fields of computing or electronic engineering. According to the TRECVid guidelines "The search task models that of an intelligence analyst or analogous worker, who is looking for segments of video containing persons, objects, events, locations, etc. of interest. These persons, objects, etc may be peripheral or accidental to the original subject of the video" (TRECVid-Guidelines, 2005).

The dyads were assigned prior to the first experiment and each user completed an online MBTI personality questionnaire, using the human metrics questionnaire (www.humanmetrics.com) to identify the trait type of each person. We planned to use this data to determine if the personalities of the participants, working collaboratively as dyads affected their performances.

In addition to this, dyads were also asked to complete pre- and post- experiment questionnaires. The pre-experiment questionnaires elicited information such as the age, gender, current affairs knowledge, frequency of watching TV news and familiarity with the system used. The post-experiment questionnaire elicited information as to the users' interface preferences, the usability of the system, the aspects of the systems that they liked and the aspects that they disliked (see Appendix A for both these pre and post experiment questionnaires). Dyads were introduced to the system prior to their tasks and given a few sample tasks to ensure that they felt comfortable using the system. Fourteen males and two females participated in this experiment, and both females were paired together (i.e. there were no male/female dyads).

Each dyad completed 12 tasks i.e. were given 12 topics or items of interest to search for, with a time-limit of ten minutes for each task. In total, 24 topics were used in the task, labeled topic 0149 to topic 0172 (see Appendix B). Each dyad completed six of these tasks on the Efficiency interface and the other six were completed on the Awareness interface. The tasks were arranged using a Latin squares design to minimise learning effect and bias. Dyads completed their 12 tasks on 2 different dates as we felt that users would be too fatigued to complete training tasks on each system, as well as all of the actual tasks, in one sitting. The experiments were recorded using a CCTV camera, for post-experiment analysis. Figure 3.5 shows the experimental set-up for these user-tests.

3.2.3 Results

Results from the post-experiment user-questionnaires showed an overwhelming preference for the Awareness version of the system, with six out of eight dyads of users preferring the Awareness version and just one dyad preferring the Efficiency system (the post-experiment questionnaire dispensed is included in Appendix A). One dyad disagreed in their preference. Not only this, but dyads performed better on the Awareness version in terms of the numbers of relevant shots found for the search topics within the time limits set (see Table

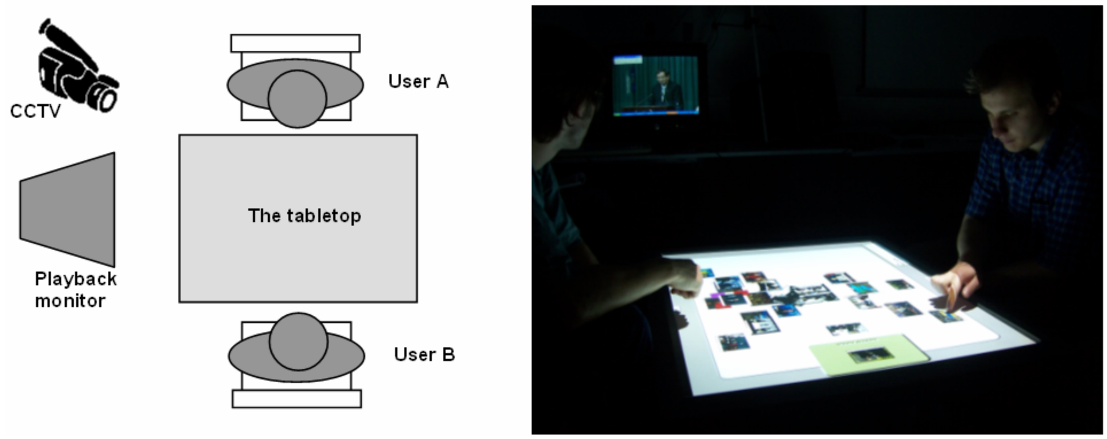


Figure 3.5: *Experiment setting for TRECVid task*

3.2, and their coordination and communication (see Table 3.3)).

In 3.2, we can see that the MAP (Mean Average Precision) score on the Awareness version was higher than that on the Efficiency version, similarly for P@10 and Recall, indicating better performance on the awareness interface. We were surprised by this, as originally we had predicted that the Efficiency version would outperform the Awareness version due to the increased speed with which the searches could be executed and results examined. The speed only of the interaction was not considered, since this out-performance would lead to the view that the efficiency version was far more interactive than the awareness version. We can also see that the overall communication of the Awareness interface was higher, with more requests and responses, and fewer coordination errors and comments. This was unsurprising as it appears logical that users would have fewer coordination errors when they have a greater awareness of

System Name	MAP	P@10	Recall
Físchlár-DT-*Awareness*	0.1529	0.7167	0.0685
Físchlár-DT-*Efficiency*	0.1372	0.6042	0.0673

Table 3.2: *Performance values for Awareness and Efficiency*

each others' actions. Also, it since the functions were shared on this interface, it would also appear logical that there would be more requests and responses.

Retrieval performance was measured using Precision and Recall metrics, which are standard metrics used to evaluate search systems (Foley et al., 2005). Precision is defined here as the proportion of shots retrieved that are relevant out of all shots retrieved. This is a real decimal number between 0 and 1, with values closer to one indicating better performance. To analyse the overall performance of each system, we used a metric called Mean Average Precision (MAP) which measures the mean of the average precisions calculated for all topics across all pairs.

Another metric, P@10 was also used which measures the Precision at the 10th document saved. Recall refers to the number of relevant shots retrieved out of all relevant shots in the collection. Again, this is a real decimal number between 0 and 1, with values closer to one indicating better performance. Table 3.2 (Smeaton et al., 2007) shows the values obtained for each of these metrics, for our Awareness and Efficiency systems. While Precision and Recall are indicators of the quality of search output and are the primary measures of success in system-oriented Information Retrieval, they are not on their own indicators of overall user interaction and design success.

Hence, in addition to looking at preference and performance data, we also manually analysed the CCTV video data gathered to examine the collaboration between the dyads. We annotated the videos along the following criteria: requests (verbal, gestural or both), responses (verbal, gestural or both), comments or coordination errors (e.g. one user playing over their partner's video, clashing hands, enlarging an image so that it interrupts the other person's work). In this case, the results were as predicted; users communicated more with each other

	Awareness	**Efficiency**
Request	563	497
Response	520	441
Coordination Error	608	702
Comment	62	94
Total	**1,753**	**1,734**

Table 3.3: *Communication recorded for Awareness and Efficiency Systems*

when using the Awareness system and fewer coordination errors were recorded with this version.

We also noted that communication between the searchers did not decay throughout the duration of the tasks.

3.2.4 Personality

In addition to looking at user preference (post-experiment questionnaire responses) and performance data, we also attempted to correlate these results with the MBTI personality data gathered from each individual to see if the performance of each dyad was in some way related to their personalities, in particular, to their group personality compatibility. Our assumption here was that dyads whose members' personalities were highly compatible, would work better together. Firstly, we performed a compatibility test on each pair to see how well-matched each dyad was. We then compared the results of these tests to the performance results of the respective pairs.

Surprisingly, we found that performance was negatively correlated to pair matching i.e. better-matched pairs performed worse on the tasks than those that were less well-matched (see Table 3.4). A possible reason for this was that pairs that were less well-matched may have been more intent on completing the task-at-hand, as they had to work harder at their coordination.

However, one must note that the collaborative activity was only for a short period of time. Possibly, for a long-term collaborative task, these results could be different. However, since time restrictions did not allow us to observe such

Pair	PersonMatch Ranking
A	4
B	8
C	=6
D	5
E	=6
F	=2
G	=2
H	1

Table 3.4: *Rank Positions for Search Performance and for PersonMatch*

a task over a longer time period, this was out of the scope of this thesis and is a possible avenue for future work.

Further details on the experiments on these two interface variations to the Físchlár-DT system can be found in (Smeaton et al., 2007).

3.3 Our Thesis Hypothesis

Having surveyed how personalities influence performance in shared collaborative tasks, and having completed a preliminary user experiment working with dyads in a collaborative video search task, we are now in a position to formulate and present the hypotheses which form the main questions to be addressed in this thesis.

In the preliminary experiments described above, the surprising results from this user-experiment led us to become interested in personality combinations of dyads working on a shared, multi-user surface interface. In particular, why did people who were apparently better-matched in terms of their personalities perform worse than those who were less well-matched? Is there a relationship between elements of an interface that users prefer and their personalities? Do dyads develop the same impression of a system, either by one user expressing delight/exasperation at certain parts of the interface or by a more timid person's opinion being altered by a more aggressive and/or outgoing partner?

In the remainder of this thesis, we attempt to answer these questions by conducting a series of five user experiments using different interfaces to fundamentally different systems. In the experiments to be reported later, we increased the number of participating dyads from eight to eighteen to strengthen the support of our hypotheses. We also conducted these experiments over a period of between two and five months (the latter was due to college holidays).

Rather than using the online MBTI test that we used for our TRECVid 2005 experiments, which had poor reliability or validity data (see Section 2.1.2), we decided to use a test that measures the *Big Five* personality traits to test the personalities of our participants. We chose the *Big Five* as it is currently the most widely accepted model of personality along five broad factors, among professionals in the field of Psychology (Pervin and John, 1999). The personality questionnaire that we chose to use for these experiments was an online questionnaire, called the IPIP-Neo ((International Personality Item Pool Representation of the NEO PI-R^{TM}) inventory and is based on the Five Factor Model of Personality.

This online questionnaire is provided by Dr. John A. Johnson, who is Professor of Psychology at Penn State University and who devised this questionnaire based on the full IPIP inventory 1,699 (Goldberg, 1981). More than 200,000 people have completed this questionnaire since it was made available on the internet and it has won an MSNBC award. There are two versions of this questionnaire available online – a 300-item questionnaire and a shorter 120-item questionnaire. The shorter questionnaire was tested on 20,000 users before being deployed on the internet, in order to ensure acceptable five-factor measurement reliability. The shorter version is the one that we used in our user-tests and generally takes between 15-25 minutes to complete (Johnson, 2008).

Given the results of the TRECVid experiments and the observations we made, we defined our hypotheses as follows:

Hypothesis 1: The personality composition of a dyad impacts the performance of that dyad, or in other words, dyads composed of certain personality types will perform tasks better than others.

This hypothesis is based on the question "why did people who were apparently better-matched in terms of their personalities perform worse than those who were less well-matched?" we posed after the TRECVid 2005 experiment above.

Hypothesis 2: Dyads with certain personality types will prefer and work better on certain interfaces.

This second hypothesis relates to the two questions "Is there a relationship between elements of an interface that users prefer and their personalities?" and "Do dyads develop the same impression of a system, (something that could potentially occur if one user expresses delight/exasperation at certain parts of the interface)" that we listed above.

Hypothesis 3: Dyads perform different tasks in a different manner and this is related to their personality.

This was not based on our previous work. However, this was something that we wished to investigate.

The remaining chapters in this thesis describe an experimental procedure that we performed, to prove/disprove the hypotheses that we have listed above. All of these experiments were performed by dyads using the DiamondTouch to test different interfaces to both competitive and collaborative systems.

3.4 Summary

In this chapter we have introduced a two-person video retrieval system called Físchlár-DT, which was developed for the DiamondTouch tabletop device using the DiamondSpin toolkit. We used this system to conduct a user-experiment consisting of eight dyads for the annual interactive TRECVid benchmarking conference series in 2005. We presented two interfaces to this system, one that focused on increasing the efficiency of individual keyframe manipulation on the tabletop and one that heightened each user's awareness of their partners' actions. The results obtained from this user study were supplied, showing that dyads preferred and performed better on the interface that heightened awareness. We also gave an overview of how the collaboration of the dyads was quantified and how this led us to the hypotheses which we will prove/disprove in the remainder of this thesis. This preliminary study provided us with a basis to our further user experimentation, the setup of which will be reported in the next chapter.

Chapter 4

Experimental Methodology

This chapter details the systems we modified and developed for use in our series of user tests. We discuss the rational behind our choice of systems for these user tests and give an in-depth description of each of the systems used. We also provide an overview of our participant recruitment and selection process and how these participants were categorised for the purpose of analysis, which follows in the next chapter.

4.1 Overview

While much research has been conducted in user interface design, taking into consideration peoples' personalities (e.g. Saati et al (Saati et al., 2005), Reeves and Nass (Reeves and Nass, 1996), Karsvall (Karsvall, 2002)), such research has been solely based on a single user working on a single computer. The research we have conducted here has been for two people, working on a single-display, collaborative computing device, namely the DiamondTouch. From our studies, we can draw conclusions, which future collaborative interface designers can learn from.

From the previous chapter, we listed three hypotheses which we aimed to prove or disprove through our user-experimentation. Since these hypotheses are very general, we propose a total of sixteen of sub-hypotheses questions, the answers to which our users-experiments and subsequent analysis would provide. These sub-hypothesis questions are as follows:

4.1.1 Hypothesis 1

The personality composition of a dyad impacts the performance of that dyad, or in other words, dyads composed of certain personality types will perform tasks better than others

What this proposes is that the personality types forming a dyad have a direct and measurable impact on their performance of the task at hand. This means that we should be able to measure a difference in performance and correlate that in some way with the personality composition of the dyad. Specific questions we can ask are:

Q 1. Do we simply focus on Extraversion as the sole personality factor to correlate to performance or interaction style of dyads ?

Q 2. Do the remaining "Big Five" personality traits affect the performances of dyads ?

Q 3. Do dyads that are more similar in terms of their personality composition outperform dyads containing very different personality types ?

Q 4. Is the interaction recorded among dyads related to their personality composition ?

4.1.2 Hypothesis 2

Dyads with certain personality types will prefer and work better on certain interfaces

This hypothesis proposes that certain combinations of personality types within a dyad will show a marked preference for certain interface characteristics. Specific questions we can ask are:

Q 5. Do individuals within dyads develop a similar impression of a system ?

Q 6. Do users prefer interfaces which model their personality along the *Extraversion* trait ?

Q 7. Do dyads perform better on an interface variant/under a task constraint variant that they like better when give two variants ?

Q 8. Is there a relationship between a user's stated opinions on a system and their interaction data ?

4.1.3 Hypothesis 3

Dyads perform different tasks in a different manner and this is related to their personality

This final hypothesis statement is a little more subtle than the previous ones and sets out to investigate how the personality combinations making up dyads, affect the interaction as well as performance of the dyads in collaborative tasks. Specific questions we can ask related to this hypothesis are:

Q 9. How does imposing different constraints on a collaborative task affect the performances of the dyads ?

Q 10. Are there more interaction instances in a collaborative version of a game as opposed to a competitive version ?

Q 11. Does the amount of interaction among a dyad relate to the performance of that dyad in our collaborative tasks ?

Q 12. Do dyads coordinate their actions well on our collaborative search tasks and is this related to their personality type ?

Q 13. Do the same territoriality tendencies exist regardless of the task or are there cases of some tasks where territoriality is irrelevant in both our competitive and collaborative tasks ?

Q 14. Do dyads with certain personalities employ different territoriality techniques than others when performing all of our tasks ?

In addition to the above questions, which are directly related to our hypotheses, there are other questions that we should investigate, which are related to the validity of our data. Specifically we should ask:

Q 15. Does performance of dyads vary to a greater or lesser extent across the different collaborative tasks used ?

Q 16. How much variability is there in the interaction among dyads across the different collaborative tasks used ?

Based on these hypotheses and sub-hypotheses questions, we devised a set of five tasks, which we also categorised in Chapter 3. In the next section, we will describe the experimental methodology and each of the five systems that were chosen for our series of user-experiments. In addition, we also show how each user-experiment provides us with or contributes to providing the answers to each of these sub-hypothesis questions.

There are many possibilities for conducting user-experiments (Castillo and Hartson, 2007), which can be broken up into two categories – remote user-testing and traditional laboratory style user-testing. Remote-user testing includes the following:

User-reported critical incident method – where users are trained to identify critical incidents and report information about them when they occur (Hartson and Castillo, 1998).

Instrumented or automated data collection for remote evaluation, where users interaction is logged by embedded metering code;

Remote questionnaire or survey - where questionnaires are sent by e-mail or are automatically triggered to pop-up when users carry out certain actions (e.g. Alertus®, 2008);

Video-conferencing supported evaluation - where evaluators are remotely connected to a usability lab via video-conferencing software (Bly et al., 1993);

Third-party lab-based usability testing - where third-party evaluators examine the design and software without recruiting users (O'Brien, 2008);

Third-party usability testing - where outside evaluators can conduct a traditional lab-based experiment at their location (Fhios, 2008).

We chose traditional lab-based user-experiments, as well as deploying user questionnaires and recording users actions via recording the interaction by CCTV camera and recording their touch-points on the tabletop's surface. Lab-based experiments are expensive in terms of both the amount of time and effort required of the evaluator. In addition, users cannot use the technology according to their typical daily habits (the first three remote usability testing methods above allowed this); it was the most appropriate method in our situation for the following reasons:

1. The DiamondTouch was only provided to our group in the University on loan; hence we had to keep the equipment in a secure lab and so could not deploy it in a general location.

2. Lab-based experiments are suited to predefined tasks (which we employed).

3. While traditional lab-based data gathering is expensive from a time, effort and financial point of view, the data that can be potentially gathered is very and often more useful in comparison to other methods. By this

> we mean that the quality of usability data in discovering, analysing and repairing usability problems which can be gathered by directly observing the interaction is much more effective than, for example, remote logging of actions, since the evaluator can see at first hand, what exactly the user is doing, where problems are encountered, their reactions and how they interact (Castillo and Hartson, 2007).

A more detailed discussion of the strengths and weaknesses of each of these testing methods is available in (Castillo and Hartson, 2007). Each system used had a distinct objective in attempting to determine a person's preference in a number of competitive/collaborative settings. The order in which the changes to these systems were presented to dyads was swapped from one dyad to the next. This was in order to see if using either system first affected their preference and/or performance i.e to account for user-bias. Hence, for each system, half of the dyads used one interface/rule first and the remaining dyads used the other interface/rule first. These interfaces and rules are detailed in the description of each system below. Section 4.3 will focus on the participants of these user-experiments; how they were recruited, selected and categorised.

4.2 Systems Studied

For our user-experiments, we recruited 36 people (18 dyads [1]) to complete five tasks on five different systems; one task per system. The participants ranged in age from 17 to 30 years, with more people falling into the 20-24 age grouping (16 participants). 33 were male and 3 were female, while 26 were undergraduate students and 10 were postgraduate students. 8 participants were studying in the field of Science and Health, 22 were in Computing and Engineering, 4 were studying Business and 2 were doing a course in the area of Law and Government. 25 people had never used a tabletop computer, 8 had used one once and 3 had

[1] The terms pairs, dyads all refer to the pairs of users participating in the user-experiments. These terms are used interchangeably throughout.

used one a few times before. 11 participants strongly agreed that their course involved working in groups, 14 quite agreed, 7 agreed a little, 1 was Neutral, 1 quite disagreed and 2 strongly disagreed. These participants were selected as they had the widest range of levels of the Extraversion personality trait, which we initially focused on (see Section 4.3).

Each system was designed and developed for use by two people sitting at opposite sides of the DiamondTouch tabletop (specifically, the sides that were longest). There was a minimum of one week between each task. In some cases, this period of time between tasks was longer, as some of the users had college breaks and exams. Each user-experiment was recorded using a CCTV camera, a procedure that had received ethical approval from the university. This was in turn used to analyse the interaction of dyads - specifically, we counted the number of requests each participant had, the number of responses they received, the number of comments each person made and the number of errors in coordination each incurred. These are explained in greater detail in Chapter 5. Users were told that they would be completing a set of five experiments, with a total duration of five to six hours over a period of time from December 2006 to March 2007. They were told that the purpose of the study was to see what kinds of interfaces to these tasks they preferred and that they would have to complete personality questionnaires. Further details on the recruitment process for these participants is supplied in Section 4.3. Figure 4.1 illustrates the setting in which dyads completed the user-experiments.

The procedure for each user-experiment was as follows. At the very first experiment, users were introduced to the DiamondTouch. It was explained to them how the technology worked and they then had the opportunity to ask questions or clarify any ambiguities that they had. Participants then read Plain Language Statements, signed consent forms and were informed that they could withdraw from the study at any stage. Following this and for all of the remaining experiments, users were asked to complete a pre-experiment questionnaire. This posed questions that recorded information about each participant's age,

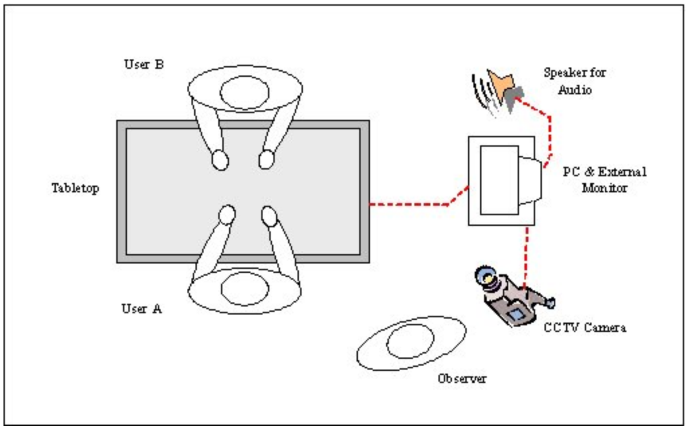


Figure 4.1: *Experiment Setup.*

background and how regularly their respective courses required them to collaborate (see Appendix A for all of the pre-experiment questionnaires used). The reason for issuing these questionnaires at the beginning of each user-experiment, was in case this data changed over the number of months that the experiments took place (i.e. perhaps they changed course during the semester, or their age changed, putting them in a different category). The questions posed also included a number of 7-point Likert scale type questions (this is also true of the post-experiment questionnaire that participants would later complete).

After some prior training, the users then completed an actual task together using the DiamondTouch. Each task required them to either use two system interface variations, or to do a task with two different constraints imposed on them. These tasks included both simple competitive games and collaborative search tasks. Our reasons for including competitive systems in this study were to determine whether personality had a significant impact on each user's performance and interaction across both collaborative and competitive tasks, and to see if this impact was similar in nature for both these kinds of tasks (i.e. do the same personality traits have an effect, or do certain traits have an effect when playing competitive games and others have an effect when performing collaborative search tasks). Once they had finished their task, participants were given a post-experiment questionnaire to complete, which elicited information about their personal preferences and opinions of the system variants they had just used (see Appendix A for the post-experiment questionnaires used). Figure 4.2 shows the layout of the five user-experiments (or sessions) we conducted.

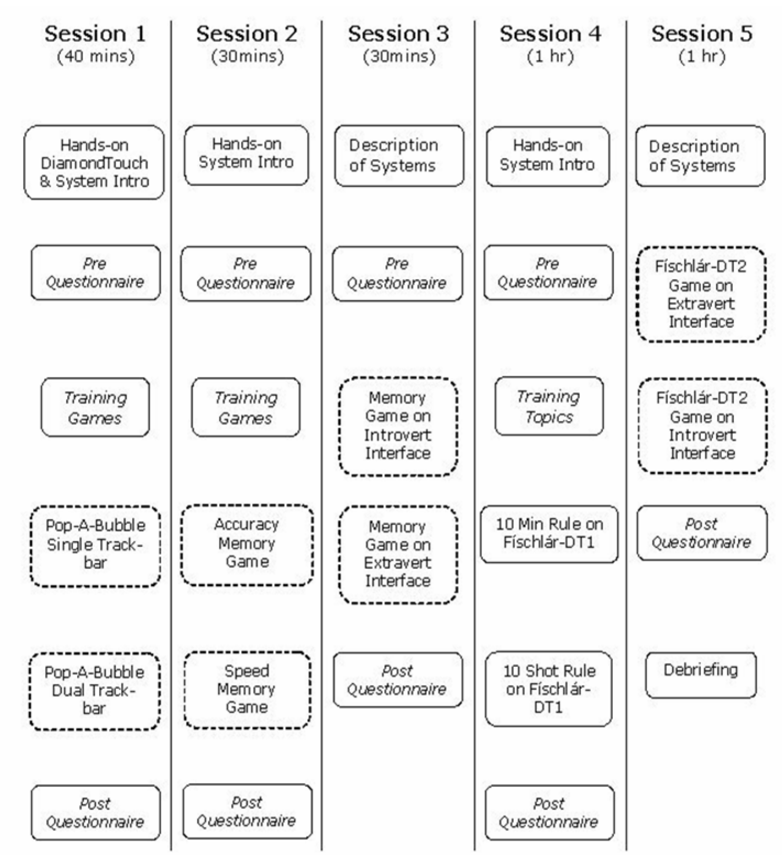


Figure 4.2: *Layout of user-experiments. The order of tasks written in boxes with broken lines may be swapped.*

Users touch-points on the tabletop, their game scores (or performance), their experiment questionnaire responses and their personality questionnaire responses were all recorded. We will now describe each of the systems used.

4.2.1 Pop-A-Bubble

Pop-A-Bubble is a competitive multi-user game, created in Visual Basic. Each player is assigned a colour, which is determined by the mat that player sits on. The object of the game is to obtain the highest score of popped bubbles. Players score 1 point for popping a bubble of their own colour, 2 points for a purple bubble and for every other colour bubble they pop, they are deducted 1 point. A specific sound is also associated with each player when they pop a bubble that increases their score. Another sound signals to the player that they have hit a bubble they shouldn't have and hence have decreased their score by one point. This application can allow up to four players. However, our user studies observed and recorded dyads using the system. Figure 4.3 illustrates this interface.

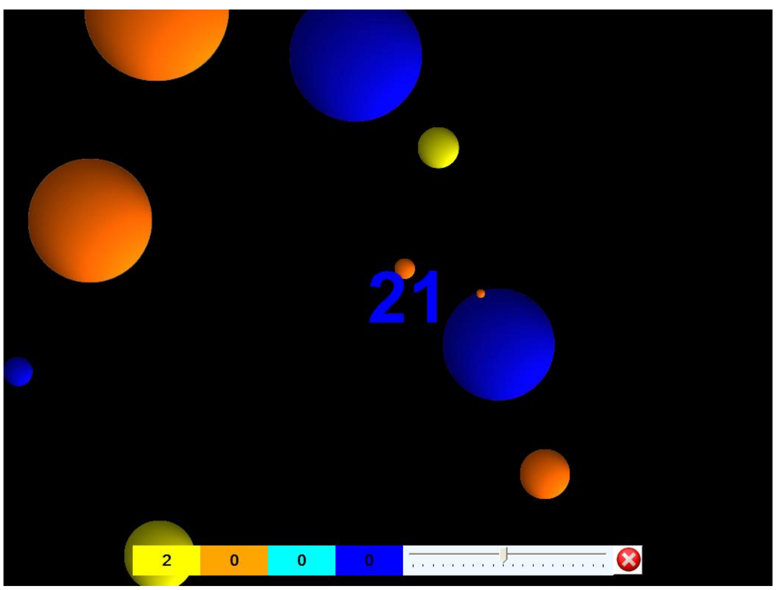


Figure 4.3: *Pop-A-Bubble Single Score interface*

This game was chosen as the first application that participants should use. There were three reasons for this.

1. It is a simple application, so participants would not be overwhelmed by using both a new technology and a complicated application (which could

have led to cognitive overload).

2. It requires immediate interaction or the player will lose. Even shy players could be seen to interact from a very early stage. As the game is highly competitive, it immediately causes people to forget their own inhibitions or fear of using a new technology, becoming enthralled with the game.

3. Using a simple game like this as the starting task also made participants familiar with the procedure of such a user-experiment and what to expect from subsequent user-experiments.

In addition, this user-experiment also aids us in answering sub-hypothesis questions 1, 2, 4, 5, 8, 12, 13, 14, which deal with the impact of personality traits on the performances of dyads, the preferences of dyad members, the coordination policy employed and the territoriality tendencies of each dyad member.

There were two interfaces to this system. One interface, which was the original Pop-A-Bubble interface, had just one score bar oriented to one player. This bar also had a track-bar, which allowed this player to change the speed or rate at which the bubbles appeared. The other interface had two such bars, each oriented to the player closest to their placement. In the single track/score-bar system, each participant played six games sitting at the side that the track-bar was, the other playing at the opposite side. They then switched sides. The side players played at in the dual track/score-bar version was irrelevant and so we felt that users did not need to switch sides. Figure 4.4 displays the second interface.

We note that these are rather discreet differences between system variants (some participants didn't even make use of the score bar or track-bar!!), but some dyads did prove to be rather opinionated in their preferences. The subtle difference between the interfaces also complemented the idea of avoiding the cognitive overload of using a new system, along with learning two new interfaces.

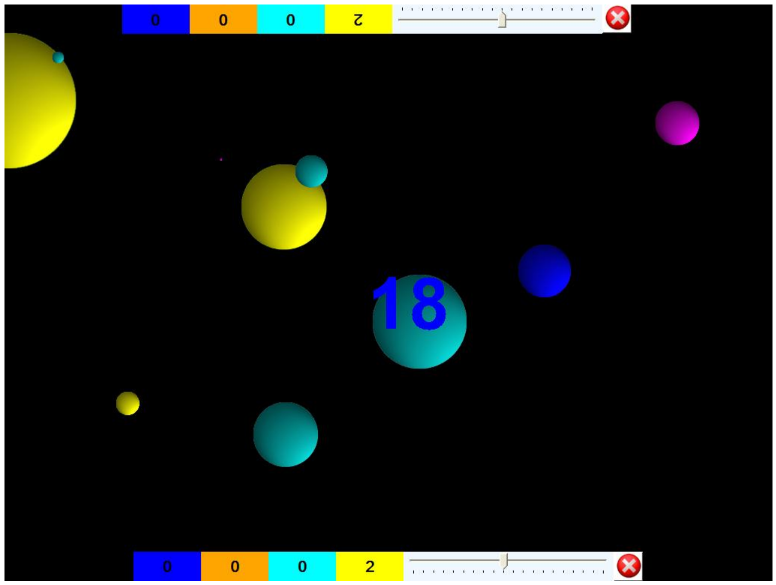


Figure 4.4: *Pop-A-Bubble Dual Score interface*

During the user-experiments for this system, we discovered that three of our users were in fact colour-blind, an issue that was not presented during prior user testing. Previous game participants (members of the CDVP who played this game before the user-experiments) did not bring the issue of colour-blindness to light, though the colours red and green had been avoided (for this very reason). However, we found that these users were still at a disadvantage as their colour-blindness extended beyond the red-green form.

The first of these users reported difficulties in distinguishing between the colours blue and purple. Since the purple bubble was worth 2 points, an alteration had to be made to the system to cater for this so that colour-blind players would not be punished for hitting a blue bubble, when they thought it was a purple one. Instead of the purple bubble being worth 2 points, we chose the cyan bubble to be worth 2 points, cyan being a safe colour for colour-blind people. They also found that the orange bubble, which was their assigned colour when playing at one side of the table, was quite difficult to identify. Therefore, we told these users they didn't need to switch sides, so the colour-blind user was always yellow.

4.2.2 Memory Game

Memory Game is a competitive card game where players try to find matching pairs of cards. A version of this was implemented in Java on the DiamondTouch, where 24 cards are displayed face-down on the tabletop. Each player overturns two cards when it is their turn. If the two cards match, the cards are silvered over, the player is given one point and another turn. If the cards do not match, control passes to the other player and this person must then select two cards. Scores are displayed at the side of the interface, so that each player can see how well both they themselves and their opponent are playing. The player with the highest number of matching pairs wins the game.

In our version, there are four sets of cards:

1. The first has pictures of different types of fruit
2. The second are a set of information signs
3. The third set contains a subset of a standard deck of cards (aces, sevens, jacks and queens, from all suits)
4. The final set of cards comprises two sets of the first twelve cards from the suit of hearts

For these user-experiments, two variations to this game were made, giving two separate tasks. The first was a collaborative version of the Memory Game, where users were required to find pairs together. This was simply called "Collaborative Memory". Two different rules or constraints were imposed to this version of the system, one which required dyads to complete the game as accurately as possible and the other which required users to complete the game as quickly as possible. These variations are detailed in the next section below. We included this system in our experiments as we believed it would be interesting to see how users collaborated in this system and whether certain people worked better and/or preferred to work collaboratively under certain task constraints (i.e. accuracy of task completion or speed of task completion). Since it is a simple system, collaboration could be analysed at a very basic and fundamental level.

Two different interfaces were built to the standard competitive version of the Memory Game (described below). These two interfaces functionally and visibly exploited both strong and weak levels of the *Extraversion* personality trait, based on research conducted by Reeves and Nass, 1996, which we described in Chapter 3 in terms of the Desert Survival Problem constructed by the authors and completed by user-experiment participants. We recall that in this study, participants preferred, perceived that they worked better on and could identify with computer interfaces that reflected their level of extraversion. High and low levels of the extraversion trait were portrayed on the interfaces in terms of the language style used (assertions vs suggestions), the initiation of interaction (i.e. computer-initiated vs user-initiated), the confidence rating supplied (high confidence vs low confidence) and the names given to the each of the interfaces. By using this version of the system in our experiments, we wanted to determine whether people who were more extraverted in personality, liked the Extravert interface better and similarly, whether more introverted people preferred the Introvert interface. If so, this would have design implications for such a tabletop competitive game.

Collaborative Memory

As stated in 4.2.2 above, Collaborative Memory is a variation of the standard Memory Game, where players must find matching pairs together. We felt that this would be an interesting alteration to make to the original game and would also exploit the collaborative and cooperative aspect of the DiamondTouch. After all, one of the key benefits of the DiamondTouch technology, is that it supports group collaboration in a more natural and intuitive way.

The game commences when the first player touches the tabletop screen.

This player is labeled Player 1 and is the player that selects the first card to overturn in this and in subsequent games. Player 2 must then select a card that he/she believe matches the already over-turned card. Their choice of card is predominantly the result of both players discussing options and sharing information about where each thinks the matching card is. Ultimately though, it is Player 2 that must decide which card to choose.

Two rules were imposed on participants in this setting. The general layout of the interface remained the same for both rules. The first rule required that users completed the game with as few mismatched pairs as possible. They were told that they could take as long as they needed to complete each game - time was irrelevant. We called this "Accuracy Memory". The other rule stated that players had to finish the game as quickly as possible, regardless of the number of mismatched pairs they uncovered. We labeled this "Speed Memory". While only accuracy scores were displayed in the accuracy version and times displayed in the speed version, we recorded both speed and the accuracy for both versions of the game, in order to see which rule resulted in the best overall performance for each dyad.

In the "Accuracy Memory" version, the number of mismatched pairs as well as the number of matched pairs were displayed along the edge of the table, at both sides of the table, so that each player could see the score (see Figure 4.5).

In the "Speed Memory" version, the time taken to complete the game they had just finished, along with the fastest time they had completed a game in, was displayed. Showing the fastest time encouraged dyads to beat this time in their current and subsequent games. Figure 4.6 illustrates this.

Each dyad played Collaborative Memory with the four sets of cards, described in Section 4.2.2 above, for each rule. This resulted in a total of eight games played. After the second game was played on each rule, players were asked to switch sides, so that each got a turn at being match-chooser and match-finder for each rule imposed. This made the task more even and fair.

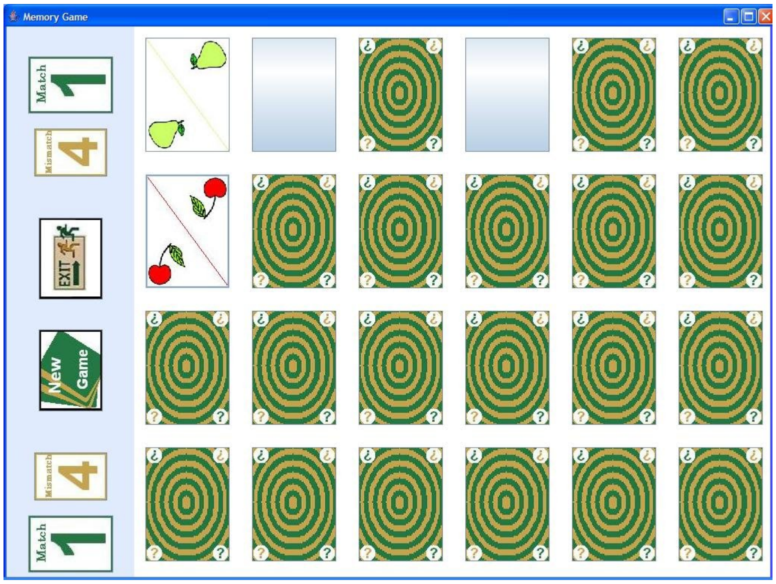


Figure 4.5: *Accuracy Memory Game interface*

For example, one could say that Player 1's task held less responsibility than Player 2's, particularly in the Accuracy Memory game, since Player 2 had to try to choose the correct matching card.

This user-experiment aids us in answering sub-hypothesis questions 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 13, 14, 15 and 16, which deal with the impact of personality traits on the performances of dyads, the preferences of dyad members, the effect of different task constraints on dyad performance and preferences, the effect of dyad preferences on their performance, the territoriality tendencies of each dyad member and the variability in dyad performances, interactions and preferences.

In the next section, we will describe the competitive version of this game and the different interfaces we designed in order to test the dyads' interface preferences.

Competitive Memory Game

This is really the original Memory Game described above in Section 4.2.2. For our experiments, this was a two-player game and both an "Extravert" interface

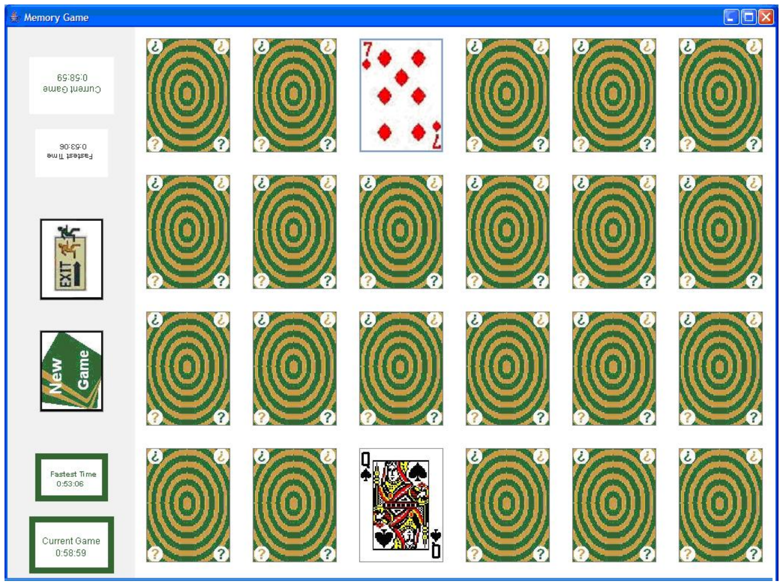


Figure 4.6: *Speed Memory Game interface*

and an “Introvert” interface were designed. The Extravert interface is shown in Figure 4.7. We chose highly saturated colours and straight lines where possible, as this was in line with previous research conducted by Pickford (Pickford, 1972). For our Introvert interface we chose much paler, more pastel colours and used rounded text and lines where possible. This resulted in a much less imposing and a much more subtle interface. This is visible in Figure 4.8

In addition to the differences in the aesthetic appearance of the interfaces, we added further extraverted “behaviour” in the form of a “Hints” functionality. This builds on the research conducted by Reeves and Nass (Reeves and Nass, 1996), where extraverted people were observed to prefer interfaces that initiated interaction between themselves and the computer device, while Introverted people preferred interaction that they initiated themselves.

In our Extravert interface, selecting a card caused a number of other cards to become highlighted – each card became surrounded by a bright pinkish-red border. These were cards that the computer believed contained the match to the selected card. The highlighting was also accompanied by a confidence rating, displayed at the side of the interface. This rating was a measure of how

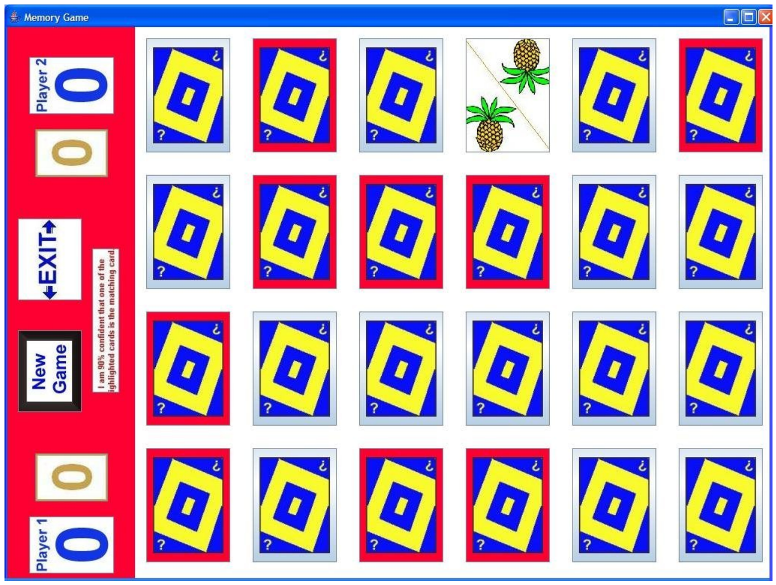


Figure 4.7: *Extravert Memory Game interface*

confident the computer was that the matching card was one of the highlighted cards. This was a rating that varied randomly between 85% and 100%, so it was always very confident. This, we felt, reflected the personality of extraverted people, who are generally very confident. It is also important to note that the card matching the selected card was always one of the highlighted cards. This is also shown in Figure 4.7.

In contrast, the Introvert interface contained a “Hints” button. Here, the user had to click the button in order to discover what cards the system would suggest. This was in keeping with the user-initiated action preferred by introverted people. The associated confidence rating was also much lower, randomly ranging from 50% to 65%. This was again in keeping with the natural traits of introverted people, who are less confident than their extraverted counterparts. This can be seen in Figure 4.9.

Once again, each dyad played Competitive Memory with the four sets of cards for each interface, totalling eight games played. The players swapped places when they had finished their four games on the first interface. This was so that each got a turn at being Player 1 for the first game. The player who

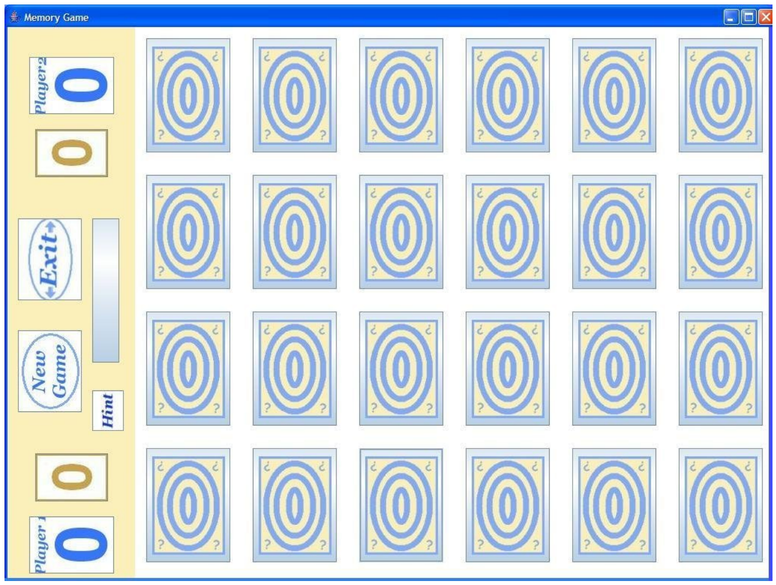


Figure 4.8: *Introvert Memory Game interface*

found the last matching pair at the end of a game, started the next game; hence there was no need for players to switch places after two games on each interface. Also, scores were allocated to each player at the side they started the game sitting at. Switching sides would have made no sense in determining the winning player.

This user-experiment aids us in answering sub-hypothesis questions 1, 2, 4, 5, 6, 8, 10, 13 and 14, which deal with the impact of personality traits on the performances of dyads, the preferences of dyad members, the coordination policy employed and the territoriality tendencies of each dyad member.

The next section describes the Físchlár-DT system and the alterations that were made to it.

4.2.3 Físchlár-DT

Físchlár-DT is a two-person video search system that we, in the Centre for Digital Video Processing (CDVP), built on the DiamondTouch for the TRECVid (Text Retrieval Conference for Video) 2005 workshop (Foley et al., 2005). A

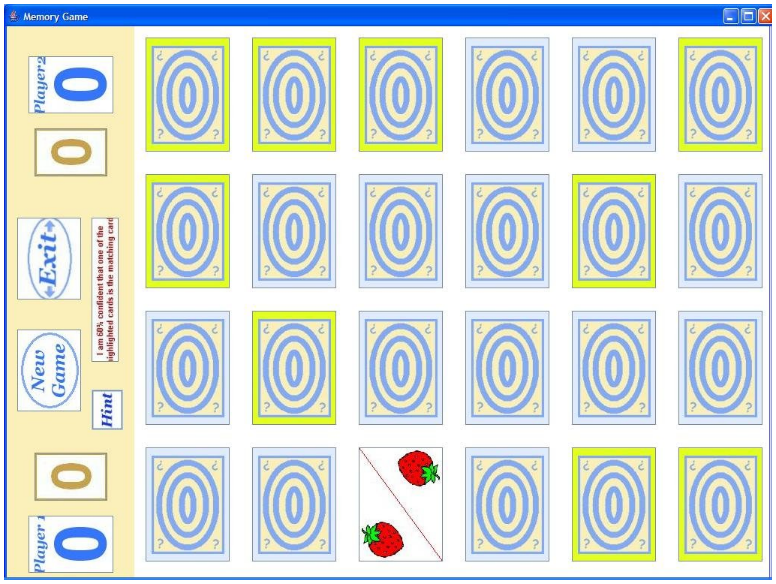


Figure 4.9: *Introvert Memory Game with cards highlighted*

more detailed description of the TRECVid activity was provided in the previous chapter.

We selected the TRECVid approach to the evaluation of our Físchlár-DT system variants for the following reasons:

1. It offers an "unrivalled series of direct performance comparisons of retrieval techniques" (Sparck Jones and Willett, 1997).

2. Its test collection is limited in size (although the collection is still large) and pooled relevance judgements are available, so that retrieval performance is easy to measure.

3. The TRECVid test collection has grown in terms of volume and diversity of the videos indexed i.e. in 2003-2004 the test collection consisted of mono-lingual tv news, in 2005-2006 the collection was broadened to include 80 hours of tri-lingual tv news (chinese, english, arabic), thereby providing a substantially large, broad and well-defined collection.

4. Our in-house retrieval engine enabled various functions pertinent to such

a video retrieval task (such as “Find Similar”, “Browse”, “Play” and “Save”) to be used.

5. Topics were broad enough that it was possible to allow for each user’s own interpretation of relevance.

A repository of 80 hours of news broadcasts from a number sources, which included NBC, CNN, as well as Chinese and Arabic news networks, was supplied by NIST for the 2005 video retrieval task. The aim of this search task was to find as many video shots as possible that were relevant to a given multimedia topic, using a video search system over the supplied repository. An example of this would be to “Find shots of Tony Blair”. 24 topics, numbered 149 to 172, each containing a text description, some sample images and some sample shots, were supplied for TRECVid 2005 (Smeaton et al., 2005). Each shot was represented by an image called a keyframe (recall that a keyframe is a representative image from a single shot of video).

Físchlár-DT was developed using the DiamondSpin SDK (Shen et al., 2004), to easily handle the rotation and orientation of objects on the interface (see Chapter 3 for more details on the DiamondTouch and DiamondSpin SDK). Our system was the first collaborative search system built for this video search benchmarking activity. As previously stated, we built two interfaces to the system for the TRECVid activity; one which focused on creating and maintaining high levels of user-awareness and the other, which enabled users to work more efficiently. The awareness interface had a number of drag-to spots or hot-spots, where the user could carry out a number of actions. The efficiency version used pop-up menus to invoke these functions and hence allowed users to work more individually and discretely. Here, we describe only the awareness interface as this was the one we used for the user-experiments dealt with in this thesis.

We chose the awareness version of Físchlár-DT over the efficiency one for a number of reasons. Firstly, post-analysis of our TRECVid interactive experiment indicated that a high-awareness facilitating interface was a more suitable interface for collaboration (Smeaton et al., 2007). Dyads could visibly see each other dragging keyframes over to certain areas, which enabled them to react accordingly. This allowed for better coordination of actions and contributed to a higher level of team-spirit, which was supported by a smaller number of coordination errors recorded for dyads using the awareness interface. The issue of awareness in the design of collaborative systems has also been the focus of much research, which re-enforces its importance in such systems. Secondly, it was the most preferred interface of participants in our TRECVid 2005 interactive experiment (Smeaton et al., 2007).

With the awareness interface, the user can type a text query, using a pop-up keyboard, into a search box located in the bottom right-hand corner of the screen. The "search" button can then be pressed and up to 20 keyframes displayed around the table. The more relevant the shot, the closer to the centre of the table is the keyframe's placement. The awareness interface has a number of drag-to spots or hot-spots, where the user can carry out a number of actions. This is illustrated in Figure 4.10.

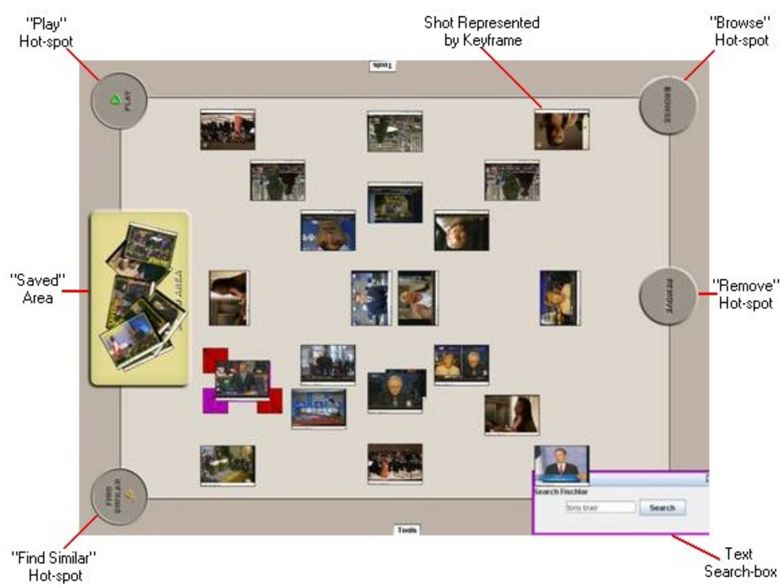


Figure 4.10: *Físchlár-DT: Awareness interface*

Each hot-spot has associated with it a distinctive sound, that can be heard when a user drags an image on top of that particular hot-spot. Not only does this let the other user know that a specific function has been invoked, thus increasing awareness, but it is also a form of feedback to the user, that the function was invoked properly.

The user can drag a keyframe onto the “Play” hot-spot, which opens up the shot it represents on an external monitor and plays it. This is useful in determining the relevance of the shot. For example, the camera may pan to show the object or person the user is searching for, or the object may just appear at some point in the shot. The “Browse” hot-spot shows the next ten and previous ten shots, in that particular news broadcast, to the one selected. This is useful if a keyframe itself does not contain an image of what the user is looking for, but looks like it may appear in shots close to the current one.

The “Find Similar” hot-spot displays 20 keyframes that are similar to the selected keyframe, by comparing MPEG-7 descriptors of that keyframe to the rest of the keyframes in the repository. These MPEG-7 descriptors include an Edge Histogram Descriptor and a Local Colour Descriptor. “Remove” deletes the selected keyframe from the screen, not to be retrieved again for that particular topic. Finally, if a keyframe is moved into the “Saved Area”, the shot is deemed relevant and marked with a yellow border and stamp to indicate that it has been saved. Any or all of these functions can be invoked by dragging a keyframe over the relevant hot-spot. See 4.10.

Similar to the Memory Game system, one user-experiment imposed two different rules on the participants in their search task on this system. The last user-experiment required participants to use two interface variants to the system. Again, these interface variants exploited both strong and weak levels of *Extraversion*. This resulted in a variation in the levels of awareness, in order to assess our participants’ reactions (see Section 4.2.3).

Físchlár-DT 1

The interface used in the first search-based task was the awareness interface we have just described. As stated in the last section, two rules were imposed on dyads in this particular user-experiment. The first imposed a ten minute time-limit, where dyad's had to find as many shots as possible that were relevant to a specified topic. Two topics were selected for this rule; one topic was to "Find Shots of Condoleeza Rice" and the other was "Find Shots of People Shaking Hands". The order of these topics was switched from dyad to dyad so that there would be no bias in performance in terms of the first or last topics searched for. This counteracts differences between the groups that may have existed, which may have interacted with their performance on the search task and cause the observed result.

The other rule demanded that dyads find ten shots in total that were relevant to a particular topic. The first of these topics was to "Find Shots of Tanks or Other Military Vehicles" and the remaining topic was to "Find Shots of Banners or Signs". It was checked before imposing this rule that the two selected topics had at least ten relevant shots in the repository. The interface remained the same for both rules. We employed these two rules in order to compliment the two rules imposed on our Collaborative Memory game. From this experiment, we too would be able to determine whether people performed relatively better or preferred working under a certain task-constraint.

The topics used here were a subset of four of the 24 topics used for the TRECVid 2005 interactive video retrieval experiments. The reason for selecting these topics was that relevance judgements were supplied for these topics. Relevance judgements in this case refers to lists of shots that have been considered to be relevant to a topic by assessors who judged TRECVid submissions at NIST. This made these topics appropriate to use in this setting for judging the performance of dyads, so that the relative performance of dyads across both constraints could be compared. This was also in keeping with the TRECVid methodology, which we used for this experiment.

Figure 4.2 from earlier in this chapter, shows the procedure of the experiment. Some training topics were given to each dyad to ensure that they were comfortable with using the system, as this task was a lot more work-oriented and had a higher cognitive effort associated with it.

This user-experiment helps us to answer sub-hypotheses questions 1, 2, 3, 4, 5, 7, 8, 9, 11, 13, 14, 15 and 16, which deal with the impact of personality traits on the performances of dyads, the preferences of dyad members, the effect of different task constraints on dyad performance and preferences, the effect of dyad preferences on their performance, the territoriality tendencies of each dyad member and the variability in dyad performances, interactions and preferences.

Físchlár-DT 2

For the last user-experiment, we altered the interface to the Físchlár-DT system, to give an Extravert and Introvert interface. Figure 4.11 below shows the Extravert interface. We again chose bright, highly saturated colours, boxes and sharp edges (Karsvall, 2002). We removed the "Find Similar" hot-spot and instead, when a user saved a keyframe, the system displayed four keyframes representing shots it thought were similar to that saved. These were then displayed in a "Suggested Clips" area, located at the centre of the table. This was really an automatic invocation of the "Find Similar" function, though on a much smaller scale as only four keyframes were displayed instead of 20. We felt that this was in keeping with the idea that system-initiated interaction is preferred by extraverted people (Reeves and Nass, 1996).

The final difference in this interface was the change in location of the "Browse" hot-spot, which was moved from the top right-hand corner of the table, to the bottom left-hand corner. This was where the "Find Similar" hot-spot was originally located and we moved it to make the interface more balanced i.e one hot-spot on each side of the table and two mid-way, along the edges.

Figure 4.11: *Extravert Físchlár-DT 2 interface*

For our *introvert* interface (see Figure 4.12), we used softer, more pastel colours. The hot-spots were round in shape, the text was also more rounded and we moved the saved area to the centre of the table. All of the functions were duplicated for each user, allowing them to work more independently and quietly, which we felt was in keeping with the introverted personality type. This however, did result in users being less aware of each others' actions. This will be examined closely in the results chapter (Chapter 5), to see if it made a significant impact on the participants performance and their personal opinions. The "Find Similar" hot-spot was brought back into this version.

Dyads searched for 3 topics on each of these interfaces, totaling 6 topics altogether for the entire session. These again, were a subset of the TRECVid 2005 topics (see Appendix B: Topics 0160, 0171 and 0151 on the Extravert interface, and Topics 0156, 0150 and 0172 on the Introvert interface). There was a 5 minute limit imposed on dyads for each topic, to find as many shots as possible that were relevant to that topic. As previously stated, the order of presentation of the interfaces was changed for each dyad (see Section 4.2).

This user-experiment aids us in answering sub-hypothesis questions 1, 2, 3,

Figure 4.12: *Introvert Físchlár-DT 2 interface*

4, 5, 6, 7, 8, 11, 12, 13, 14, 15 and 16, which deal with the impact of personality traits on the performances of dyads, the preferences of dyad members, the coordination policy employed, the territoriality tendencies of each dyad member and the variability in dyad performances, interactions and preferences.

We will show in the proceeding chapter that as a result of answering our hypotheses sub-questions, achieved by the analysis of our user-experimental data gathered, that the first hypothesis we proposed (i.e. *Dyads with certain personality types will prefer and work better on certain interfaces*) is proven. However, we can only partially prove the remaining two hypotheses asserted i.e. *Dyads with certain personality types work better on certain interfaces* and *Dyads perform different tasks in a different manner.*

4.3 Participant Recruitment

Participants were recruited from the general university student population. An e-mail requesting the participation of pairs of people to complete 5 user-experiments was sent campus-wide. The e-mail also stated that respondents selected for participation would be financially rewarded, in order to attract more people. The e-mail stated that the study required them to play games and complete enjoyable teamwork tasks for 5-6 hours over a three month period (i.e. from mid-December to mid-March). The specification that pairs of users volunteer was important for this study, as it was imperative that participants express their natural personalities and tendencies as quickly as possible. Completing a task with a person one wasn't familiar with could inhibit their natural collaborative/competitive instincts. This would be particularly evident in the case of introverted users.

While it could be argued that mixing the groups could determine how people behaved when they had a partner/competitor that they did not know, we wanted to reflect a typical group-working scenario, where people work together for longer periods and so get to know each other over time. Since we had only five user-experiments, each between half an hour to one hour in duration, we believed that recruiting pairs would mean that the participants' behaviour would reflect their behaviour over a longer group-working situation. It must also be noted, that I was present in the room for each and every experiment, which may have caused people to behave differently. Having a partner that was also unfamiliar may have worsened this effect. We must also consider that participants were using a new technology. Both using a new technology and carrying out tasks with someone one wasn't familiar with may have lead to certain users becoming overwhelmed. In addition, we believed that requesting pairs of users to respond would encourage people to be more committed to completing all five experiments – each would not want to let their partner (whom they know) down by not showing up.

Respondents to this recruitment e-mail were then requested to answer six short questions from the IPIP-Neo online personality questionnaire (Johnson, 2008). These six questions appeared to prevalently measure the *Extraversion* trait, after numerous completions of this questionnaire. We requested answers to just six questions so that potential participants would be more encouraged to reply to an e-mail with a short questionnaire – a questionnaire containing 120 questions may have seemed too long, if respondents knew there was a risk that they would not be selected. The reason for measuring the strength of the *Extraversion* trait at this stage was so that people with as wide a range of levels as possible could be selected. We felt at this stage, that the *Extraversion* trait was going to be the most significant in determining the type of collaboration exhibited and the performance of the dyads.

The idea of a second e-mail requesting information about the respondents worked well as it not only allowed for selection, but it also showed a certain level of commitment by these respondents to the study. Those who really weren't particularly bothered did not respond to this second e-mail. This meant that the majority of respondents to the second e-mail that were selected, actually completed the entire set of user-experiments. However, it also meant the total number of respondents that actually took part at this time was quite small – just 26 people in total, making 13 pairs. Some months later, 10 more people volunteered and were added, which raised this number to 36 people (18 pairs).

The first batch of user-experiments began in early December 2006 and continued right through until April 2007. The second batch of experiments then ran from July 2007 to August 2007. Both batches completed the tasks using the same systems, under the same conditions; though the first batch did take significantly longer to complete. This was due to college vacations and exam study periods which ran from Christmas 2006 until the start of February 2007.

Of the 36 people that participated, just 3 were female. Although this is a small proportion of females vs. males, it does reflect the male-dominated gender distribution in this area (e.g. in the CDVP research group, there are currently 6 female members out of 48 members in total (9.8%)). Any other females that had responded to the recruitment e-mail, failed to show up at the experiments, again showing this disinterest by females. These three females took part in the experiments with male partners, so there were no female/female dyads.

Experiment Participants		
Introvert/Introvert	Introvert/Extravert	Extravert/Extravert
Dyad 3 (29%, 34%)	Dyad 2 (45%, 78%)	Dyad 1 (80%, 76%)
Dyad 5 (3%, 38%)	Dyad 6 (43%, 83%)	Dyad 4 (82%, 69%)
Dyad 13 (4%, 41%)	Dyad 9 (28%, 51%)	Dyad 7 (59%, 66%)
Dyad 15 (38%, 9%)	Dyad 12 (34%, 68%)	Dyad 8 (55%, 59%)
Dyad 16 (45%, 0%)	Dyad 14 (45%, 87%)	Dyad 10 (70%, 57%)
	Dyad 18 (43%, 92%)	Dyad 11 (53%, 68%)
		Dyad 17 (54%, 57%)

Table 4.1: *Breakdown of dyad Extraversion*

Other possible participant recruitment methods include personal visits to lectures around campus to inform people of the experiments and encourage them to participate, or placing poster advertisements in different locations around campus. However, we felt that e-mail was adequate by itself in our case, since it reached a large audience (i.e. all students) and enabled enough information to be transmitted to encourage people to participate. Since we received numerous responses to our e-mail, further recruitment methods were not required.

After the first experiment, dyads were asked to complete the short online IPIP-Neo personality questionnaire, in order to obtain a more accurate measure of their level of extraversion. Table 4.1 shows the breakdown of each pair in terms of their level of Extraversion. For the purpose of participant anonymity, the 18 dyads were be labeled Dyad 1 - Dyad 18 here. Dyads labeled in this manner throughout the remainder of this thesis, will refer to those listed in this table.

An introvert in our case refers to a person who had a 0% to 49% level of *Extraversion*, while an extravert refers to a person who had a level of *Extraversion* greater than 50%. While scores between 35% and 65% may be considered average, we felt that people just above the 50% level would be leaning towards a more extraverted personality. Likewise, people scoring below 50% were seen as leaning towards a more introverted personality and hence would show more introvert than extravert traits.

Once participants from the first batch of user-experiments completed all 5 tasks, they received their financial reward of €50, were thanked and told that they could follow up on any of the results if they so wished. The remaining volunteers were also thanked and told that they could follow up on any of the results if they so wished.

The systems above were tested prior to these experiments, so that issues could be brought to light and corrected by users. Pop-A-Bubble was a standard DT application and the original Memory Game was written by a Summer intern working on the Diamondtouch. In both cases, the systems were set up on the DiamondTouch, so that visiting researchers and members of our research group could play these games at a time convenient to them (as they did).

Since Físchlár-DT 1 was really the Físchlár-DT Awareness interface used for our TRECVid 2005 search system, our pilot testing for this set of experiments was our TRECVid user experiments. Similarly, our Físchlár-DT 2 interface variations were modified versions of the Físchlár-DT Awareness interface and so our TRECVid 2005 search system experiment was again used as our pilot test here.

In all of our above systems, we observed general HCI guidelines and usability principles. Most of them are actually results of many years of experiences more geared towards desktop single-user PC interaction. Hence, some of the principles have more relevance in tabletop design than others (i.e. not all principles and guideline items have that importance on table design).

Take one of the most important principles, "strive to be consistent" (Shneiderman's No 1 'Golden Rule', 2005) - this is very useful rule for GUI desktop applications and Web design because the interfaces usually have many different pages or sections, panels and a number of messages. When there are many such different parts, certainly being consistent amongst them becomes an important issue. However, tabletop interfaces usually do not have many such changing 'pages' or panels (indeed, we think it is not a good thing to have many different panels or pages swapping all the time on the table) - so being consistent has much less significance on table design.

On the other hand, "Speak the user's language" (Nielsen's No. 2 Heuristic, 1994) seems to have high importance on the tabletop: we want to make the table elements as simple and user-oriented as possible, metaphors such as card flipping, moving objects around, sweeping the table, etc. can be nicely used on the table.

"Provide feedback" (both Nielsen, 1994 and Shneiderman, 2005) is also relevant to tabletop interface design - as you drag an object, it immediately feeds back following the finger. For example, as in our Físchlár-DT systems, when an object is placed on a 'hotspot', it makes a distinctive noise, providing feedback that it was placed on the right location. Another rule observed in our Físchlár-DT systems was to "Provide easy reversal of actions" - if you want to unsave a shot, you simply drag it away from the hotspot, this is the support of that rule.

"Shortcuts" (Nielsen, 1994) is not very relevant to table - the priority of table will be more for obvious, simplistic, intuitive actions rather than efficient, time-saving actions. Too many shortcuts on a table can hinder collaboration (because user A will be easily able to do something without letting user B know). Shortcuts in general hinder workspace awareness.

"Prevent errors" (Nielsen, 1994) is well catered for on our systems – we want more physically-oriented actions, if a user does something that is not right, it just doesn't happen (rather than popping up an error window). Preventing user-driven errors has been taken into consideration well on the Físchlár-DT applications, and in the memory game. If it is not the user's turn, he/she can never even accidentally overturn a card – this is a prevention of error thanks to DT's multi-user recognition through seating.

4.3.1 Evaluation Metrics

In order to analyse our data, it was important that we selected the most appropriate evaluation metrics, so that meaningful and accurate results could be obtained. Standard evaluation metrics include averages and standard deviations, which we included in our analysis. For the type of data analysis that we wished to conduct, and given our experimental setup, there were a number of statistical methods that were available to us to use. These included:

1. **Regression**

   Regression Analysis is concerned with tracing the distribution of a dependent variable, or some characteristic of this distribution (e.g. its mean) as a function of one or more independent variables (Fox, 1997, p.15). There are two types - simple regression, which has a single independent variable and multiple regression, which has more than one independent variable.

2. **Analysis Of Variance (ANOVA)** One-way analysis of variance (one-way ANOVA) is used to compare the means of two or more groups (the independent variable) in one dependent variable to see if the group means are significantly different from each other. Factorial ANOVA on the other hand, used when you have one continuous (i.e. interval or ratio scaled) dependent variable and two or more categorical (i.e. nominally scaled) independent variables (Urdan, 2005, p. 101 and p. 117).

3. **t-Test**

   This compares two means to see if they are statistically different from each other. There are two types – a Dependent Samples t-Test and an Independent Samples t-Test.

4. **Spearman rank correlation**

   This is a special, non-parametric case of the Pearson product-moment rank correlation, where the data to be correlated consists of two sets of ranks, indicating the ordinal position of the subjects on each variable (Cohen et al., 2003, p. 31).

Given our experimental setup, it was found that using Regression and ANOVA were inappropriate in our analysis. Firstly, for Regression, we had numerous combinations of dependent and independent variables, so much so that the results proved to be meaningless given our small sample size. For instance, in order to identify a relationship between the performances of the dyads on different tasks (where performance is defined as games won, time-taken to complete a collaborative search task, the accuracy with which a collaborative card game was completed etc.) and their personalities, regressions had to be carried out for each task performance and a single personality trait. Then a regression had to be carried out on a combination of two personality traits with each of the task performances, then a combination of three traits etc, not to mention the different trait combination metrics that we had to use i.e. when combining both personality trait scores of dyad members for each dyad in the collaborative tasks. Conducting all of these regressions with such a small sample size, would not have led to robust results and in the worst case, would have led to misleading results.

To use ANOVA, we had to divide each of the personality trait scores into three groups: low (0-34), medium (35-65) and high (66-100). We note that in our experimental setup, we had no control / experimental groups. Problems also arose when we attempted to combine multiple personality traits. Hence, it was deemed inappropriate.

We decided then to use the t-Test and Spearman's rank correlation coefficient metrics to analyse our data. We used the t-Test to compare the means of the performances of dyads on the Speed Memory game and the Accuracy Memory game, so that we could determine if they were statistically different. We used Spearman's Rank Correlation mainly to determine statistical relationships between the personalities of the dyads and their performances and interactions in each of the tasks. This enabled us to provide more robust answers given our sample size, as ranks were used rather than ordinal values. These results and the conclusions drawn from these results are given in the next chapter.

4.4 Summary

In this chapter, we introduced and described each of the systems used and evaluated in our user-experiments, those being Pop-A-Bubble, Memory Game (Collaborative and Competitive versions), Físchlár-DT 1 and Físchlár-DT 2. In addition, we described each of the various task constraints and system interface variations that we implemented for these systems. These were selected in order to provide answers to 16 hypotheses sub-questions, which we also introduced at the start of the chapter. The answers to these sub-questions would enable us to prove or disprove the hypotheses asserted. We presented a number of different methods which we could have used to undertake our experiments and supplied reasons as to why we chose the method of laboratory user-experimentation that we did. We saw that each of these systems test a specific notion in UI design for single-display, touch-sensitive groupware technologies. We outlined the participant selection process for our user-experiments – how they were recruited and why they were selected. Finally, we gave a brief outline of the evaluation metrics that were available for us to use for our data analysis, as well as providing the reasons why we selected the t-Test and Spearman rank correlation metrics. In the next chapter, we will look at the actual analysis of the data gathered, as well as how this set of experiments enables us to answer the hypothesis sub-questions.

Chapter 5

Experimental Results

In this chapter, we provide a description of the data gathered, both automatically and manually, from the user experiments that we carried out. This includes both data gathered by the various individual systems – such as user touch-points on the display, performance of our individual users in terms of how they "scored" in the game or performed in the search task and CCTV footage – as well as data supplied by users in the form of questionnaire responses. Following this description, we then analyse and interpret this data in terms of how it answers various questions we posed, which help to validate the hypotheses that we presented earlier.

5.1 User Data Collected

Accurate analysis of the user experiments we conducted for this thesis relied upon the collection of a body of data, both directly from the users themselves, as well as implicitly by the systems they worked on and the CCTV camera used in the experimental set-up (see Chapter 4, Section 4.2). The data collection comprises of the following:

User-supplied data, which consists of:

1. Questionnaire responses from users (both pre and post-experiment questionnaires).
2. Personality Questionnaire responses from users, in terms of how each user rates on the so-called "Big Five" of personality traits.

We also have a range of **implicitly gathered data**, which consists of:

1. Performance Data in terms of how our users scored in the game experiments, as well as the number of relevant shots saved in the search tasks.
2. User touch-points on the tabletop, meaning exactly where on-screen each user's touch-points were located.
3. CCTV footage of the user experiments, which have been manually annotated to locate, among other things, the interactions of the users.

In the following subsections, we describe each of these data elements in detail, presenting graphical examples where appropriate.

5.1.1 Questionnaire responses

As previously stated in Chapter 4, participating dyads completed pre and post-experiment questionnaires. Questions posed were either in direct form, requiring either a Yes/No answer, or an essay-type response from each participant, or else they required participants to respond by choosing one point on a 7-point opinion scale in order to rate agreement with a statement.

The pre-experiment questionnaires elicited background information about each participant's age, previous experience with using tabletop technologies and previous experience in playing the particular game or completing the task at hand. The post-experiment questionnaire elicited information concerning each user's impression of the system they had just used, the aspects that they liked/disliked most and any comments that they wished to leave about the system. An example of responses from all participants to both pre and post-experiment questionnaires are included in Appendix A.

5.1.2 User touch-points on the tabletop

The points that each user touched on the tabletop were recorded for all users and for all experiments in order to determine if the issue of territoriality on the tabletop applied in all user experiments and under differing conditions. These touch-points were obtained using the *getPoint()* method supplied by the DiamondTouch SDK and referred to x/y coordinates. These x/y coordinates corresponded to the resolution of the display i.e. x coordinates ranged from 0 to 1023 and y coordinates ranged from 0 to 767, which corresponds to the screen resolution of 1024 x 768 on the DiamondTouch PC we used. By graphing these touch-points, we can visualise where exactly each user was interacting with the tabletop and whether their interaction with the tabletop was predominantly located in particular parts of the surface. An example of such a graph is shown in Figure 5.1 from the the Pop-A-Bubble system, which is non-territorial in nature. This is reflected in how the touch-points are distributed on the display.

One should note that our touch-point data for our Físchlár-DT 1 system could not be broken up into the separate rules imposed for each experiment. This was because touch-point data was recorded for the entire experiment, without breaks for each topic. However, we changed this for the Físchlár-DT 2 interfaces, so that touch-points were recorded for each interface individually.

5.1.3 CCTV Footage

Each user-experiment was recorded using a CCTV camera, which was placed at a height above the table and allowed to capture the entire interaction of

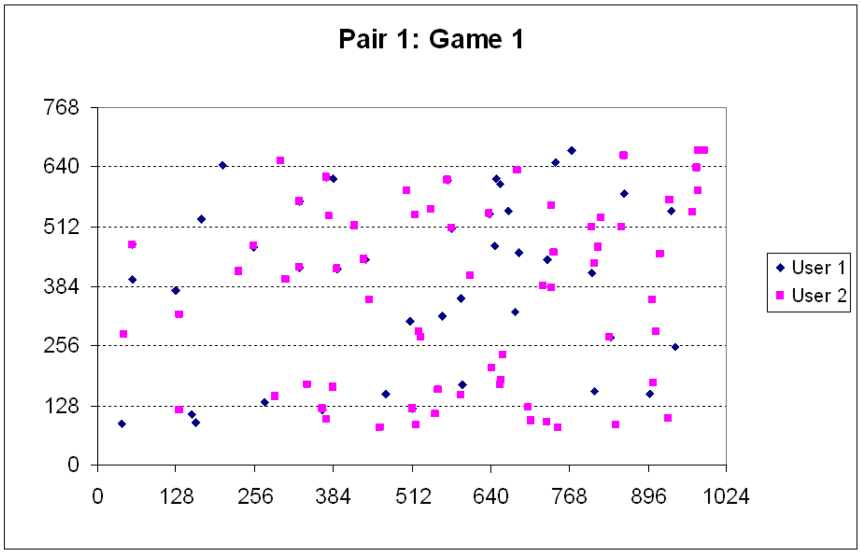


Figure 5.1: *Graph of Touch-points for a dyad using the Pop-A-Bubble system*

the dyad, both aurally and visually. Subsequently, each video recorded was manually annotated in order to create and populate a taxonomy of interaction instances, aligned with the time that these instances occurred (this taxonomy is described in more detail below). Doing this enabled the interaction between the dyads, and their overall collaboration over time, to be statistically analysed.

Taxonomy of interaction instances

Based on our observations of how dyads were interacting during the experiments, the interaction instances, as observed on CCTV capture of those experiments, were grouped into four main categories - Request, Response, Comment and Coordination Error. We also used this categorisation in our previous work discussed in Chapter 3 (Smeaton et al., 2007). We now describe each of these in turn.

Request (voice, gesture and both)

A request is deemed to be a question posed by one member of the dyad to the other. This can either be a verbal request, for example, "Can you pass me that image?"; a gestural request, for example placing an image in a queue beside a hot-spot on the tabletop, implicitly requesting the other person to place the image on the hot-spot at an appropriate time; or both verbal and gestural, which is a combination of the two.

Response (voice, gesture and both)

A response is deemed to be an answer to a question posed. This can either be a verbal response, for example, "Sure - here's the image you want."; a gestural response, for example, placing an image queued by another person onto a hot-spot; or both verbal and gestural.

Comment

A comment describes any other form of verbal utterance that is not a request or a response, for example, "Wow - this is cool!".

Coordination Error

A coordination error is an action that one person carries out that interferes with the other person's work or actions. An example of this would be if one person re-sized an image, covering the entire display or part of the work area of the other person and thereby interrupting the other person's work. Another example would be if one person invoked a function unexpectedly, without coordinating with the other user, which results in a change to the display when the other person was unaware that this was going to happen.

Figure 5.2 provides us with one example of an illustration of the number of interaction instances over time for the dyads using our Físchlár-DT 1 system (10 minute rule). This is the total interaction across both topics for this rule and for each dyad. The downward trend over time in this graph shows us that the number of dyad interactions generally decreased over time, as well as showing us the level of verbal communication and gestures that each dyad had over time i.e. those that had more communication were closer to the base of the graph.

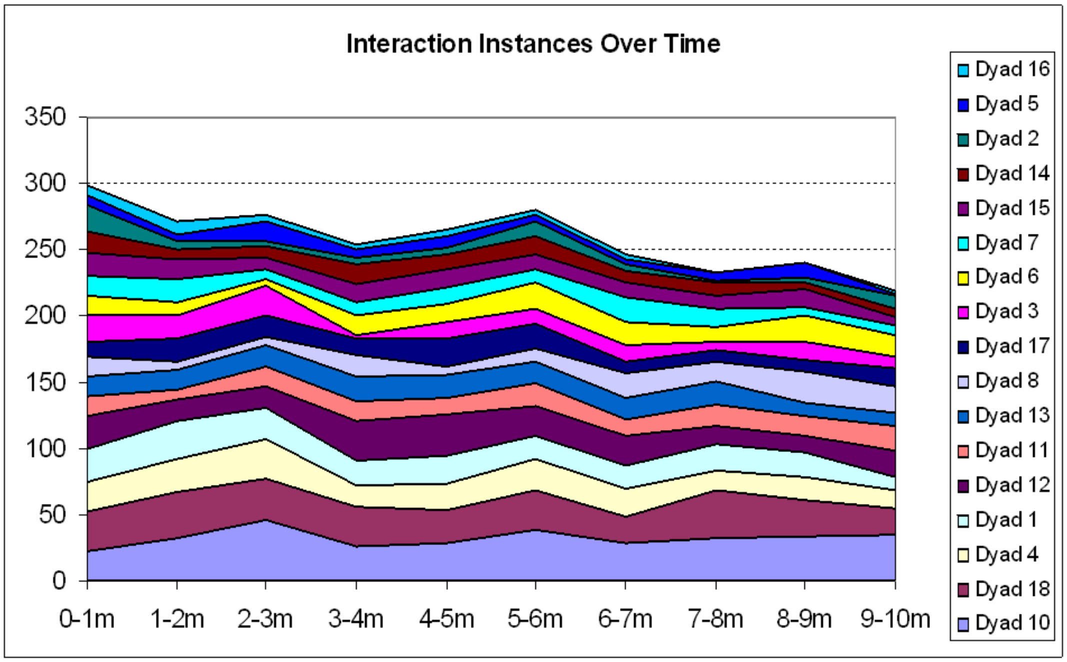


Figure 5.2: *Interaction Graph for dyads using Físchlár-DT 1*

5.1.4 Performance Data

For each user-experiment that we carried out with our dyads, we recorded each dyad's performance data, the nature of which differed from task to task. The purpose here was to find out if each dyad's performance differed when using two variations of a system and if so, which variation they performed better on.

Pop-A-Bubble: The performance data gathered for Pop-A-Bubble consisted of the scores each player achieved for each game. In addition, the number of bubbles they popped that were their own colour or an "incorrect" colour were also recorded.

Collaborative Memory Game: The same data was recorded for both Time and Accuracy versions of the Collaborative Memory Game. This data comprised of the number of mismatches each dyad scored for each game and the time it took to complete each game. The rationale here was to discover whether different constraints applied to the same task made people work in a different manner and hence caused a difference in their performance, or whether they maintained the same strategy and performance for both versions of the game.

Competitive Memory Game: The winner of the game was recorded, as well as the time it took to complete the game.

Físchlár-DT 1: Data recorded in the video search task included the shots/keyframes that each dyad saved for each of the topics on which they searched. This was in order to calculate the recall figures for each dyad for each topic. Time was also recorded for both topics on the "Find 10" rule, as the goal imposed on dyads was to find ten shots, regardless of the length of time it took.

Físchlár-DT 2: Similar to Físchlár-DT 1, the recorded performance data here comprised the shots that each dyad had saved for each topic, which we used in order to calculate the recall of each dyad, and for each topic. Recording time for this system was unnecessary as each dyad had a time limit of 5 minutes to complete each task.

5.1.5 Personality Questionnaire responses

As stated in Chapter 4, each participant completed an online personality questionnaire to measure their personality along 5 factors i.e. the Five Factor Model (Johnson, 2008). Participants were required to complete these questionnaires after their first experiment and return the results in their own time, either by e-mail or in print. Summarised results of these questionnaires for all users are included in Appendix B.

5.2 Hypotheses Restated

We recall our overall hypotheses and their respective sub-questions, that we posited in both Chapters 3 and 4. These are:

5.2.1 Hypothesis 1

The personality composition of a dyad impacts the performance of that dyad, or in other words, dyads composed of certain personality types will perform tasks better than others

Q 1. Do we simply focus on Extraversion as the sole personality factor to correlate to performance or interaction style of dyads ?

Q 2. Do the remaining "Big Five" personality traits affect the performances of dyads ?

Q 3. Do dyads that are more similar in terms of their personality composition outperform dyads containing very different personality types ?

Q 4. Is the interaction recorded among dyads related to their personality composition ?

5.2.2 Hypothesis 2

Dyads with certain personality types will prefer and work better on certain interfaces

Q 5. Do individuals within dyads develop a similar impression of a system ?

Q 6. Do users prefer interfaces which model their personality along the *Extraversion* trait ?

Q 7. Do dyads perform better on an interface variant/under a task constraint variant that they like better when give two variants ?

Q 8. Is there a relationship between a user's stated opinions on a system and their interaction data ?

5.2.3 Hypothesis 3

Dyads perform different tasks in a different manner and this is related to their personality

Q 9. How does imposing different constraints on a collaborative task affect the performances of the dyads ?

Q 10. Are there more interaction instances in a collaborative version of a game as opposed to a competitive version ?

Q 11. Does the amount of interaction among a dyad relate to the performance of that dyad in our collaborative tasks ?

Q 12. Do dyads coordinate their actions well on our collaborative search tasks and is this related to their personality type ?

Q 13. Do the same territoriality tendencies exist regardless of the task or are there cases of some tasks where territoriality is irrelevant in both our competitive and collaborative tasks ?

Q 14. Do dyads with certain personalities employ different territoriality techniques than others when performing all of our tasks ?

Q 15. Does performance of dyads vary to a greater or lesser extent across the different collaborative tasks used ?

Q 16. How much variability is there in the interaction among dyads across the different collaborative tasks used ?

These sub-questions can be visualised as relationships and associations among our gathered data, as in Figure 5.3. This figure also shows how each element of the gathered data combines to address each of our sub-questions, as numbered above. By successfully identifying all of the sub-questions enumerated here as actual links between different elements of the data we have gathered, we should be able to address all of these questions through the identification and analysis of the experimental data that determines these relationships. We can also see from this diagram that analysis of these relationships can provide answers to more than one of our sub-questions.

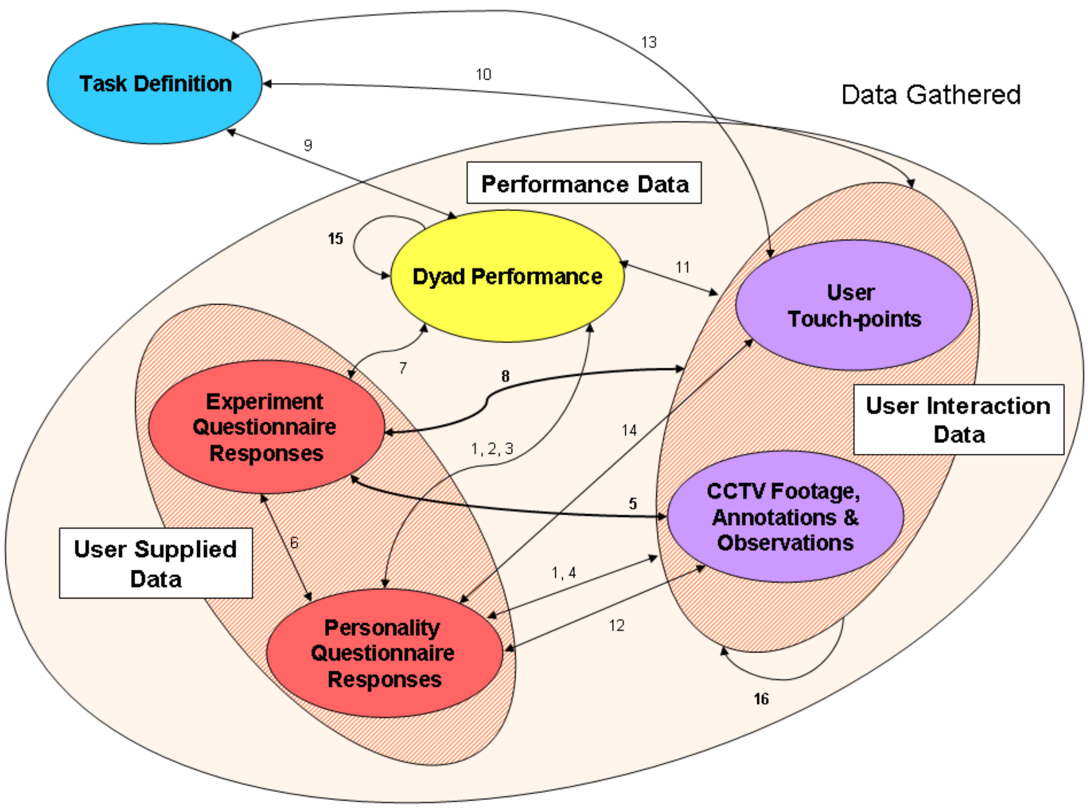


Figure 5.3: *Hypothesis sub-questions*

In the next three sections, we take each of our hypotheses and their respective sub-questions in turn, describe the aspects of the data we require to answer these questions and analyse the possible relationships that exist between the relevant data elements, in response to these questions.

5.3 Hypothesis 1

The personality composition of a dyad impacts the performance of that dyad, or in other words, dyads composed of certain personality types will perform tasks better than others

Here, we answer each of the sub-questions to this hypothesis in an attempt to prove/disprove the overall hypothesis.

Q 1. Do we simply focus on Extraversion as the sole personality factor to correlate to performance or interaction style of dyads ?

In Chapter 3, we stated that we believed that the *Extraversion* personality trait (one of the "Big Five" personality traits, Costa and McCrae, 1993) would be the most interesting and important trait to examine in these experiments, due to the highly social nature of the study. Hence, we carried out a number of tests to determine whether the combination of each dyad member's *Extraversion* personality trait had a significant relationship to the performance of the dyad. The percentage scored by each participant on the *Extraversion* trait is listed in Appendix B. We first take a look at the *Extraversion* trait and its relationship to performance and interaction style in our collaborative systems, then in our competitive systems.

Collaborative Systems

A difficulty we encountered when attempting to establish whether a relationship existed between the *Extraversion* personality trait and performance in our collaborative systems, was to try to devise a means of combining the *Extraversion* scores for both members of the dyad. We tested a number of orderings, including a metric which we called "E-Dist". This measured the absolute difference between the *Extraversion* scores of both dyad members e.g. if one dyad member scored 30% on the extraversion trait and their partner scored 50%, then their *E-Dist* would be 20%.

The idea here was that the closer each person's level of *Extraversion* was to that of their partner/opponent, the more similar and compatible they would be. Hence, we anticipated a correlation between a low *E-Dist* value and high levels of performance. To determine whether this was the case, we plotted the e-dist values for each dyad in increasing order, against their performances on each of the systems (i.e. number of mismatches in accuracy memory, time taken in speed memory, recall for Físchlár-DT 1 (10 minute rule) and both Físchlár-DT 2 variations, and time taken in Físchlár-DT 1). We proceeded then to determine if the *Extraversion* personality trait had a statistically significant on performance using the Spearman rank correlation method.

There did not appear to be a trend when we ordered dyads by ascending levels of *E-Dist* on each of our systems and plotted them on line graphs. We also ordered our participating dyads by "Average Extraversion", "Most Extravert" member and "Most Introvert" member. For the "Average Extraversion" measure, we simply took both dyad member's scores for the *Extraversion* trait and averaged them. We then sorted dyads by descending "Average Extraversion". In the case of the "Most Extravert" measure, we examined the *Extraversion* results for each member of the dyad and chose the more extravert member's result. Once we had these results (expressed as percentages), we ranked the dyads by descending order of these percentages.

We applied a similar approach to this for our "Most Introvert" member ordering. Here, we noted the member of each dyad who had the lowest percentage of extraversion and sorted all dyads by these percentages in ascending order. Again, we attempted to correlate all 3 of these orderings with dyad performance, by plotting the results on a line graph. Once more, this indicated that there was no obvious trend between any of these 3 measures and the performance of the dyads.

We then used the Spearman rank correlation method (briefly described in Chapter 4) to determine if there was a statistical relationship between each of our *Extraversion* combination metrics and the performances of the dyads. Using the Spearman rank correlation method involved creating a rank of dyads, sorted in ascending order of our *Extraversion* combination ranking from 1 - 18. Another rank was created of dyad performance. Ties in the ranks were resolved by averaging the rank of those whose *E-Dist* or performance values were the same, for example, if the 4^{th}, 5^{th} and 6^{th} ranked dyads all had the same performance value, then they would each be given a rank of 5. We then executed this correlation approach between the performance ranks of the dyads and their associated *E-Dist* ranks for each system.

The result of each of these, Spearman's correlation coefficient 'r_s', was a rational number between or equal to 1 and -1 (i.e. $-1 >= r_s <= 1$). A result of -1 indicated a negative relationship between the variables (i.e. better performing dyads will have low *E-Dist* values) and +1 indicated a positive relationship between the variables (i.e. better performing dyads would have high *E-Dist* values). A result of 0 indicated that no relationship was present between the variables.

We did this for ranks of *E-dist*, *Average Extraversion*, *Most Introvert* and *Most Extravert* on all of our collaborative systems. We noted that for Speed Memory Game and Físchlár-DT 2 (Introvert interface) the metric that showed the strongest correlation was the *Most Introvert* metric, with r_s values of -0.31 and -0.20 respectively. *Average Extraversion* showed the strongest correlation to performance for the Físchlár-DT 1 (10 minute rule) and Físchlár-DT 1 (Find 10 rule) systems, with r_s values of 0.25 and -0.44. For Accuracy Memory Game, the best extraversion combination metric was *E-Dist*, with an r_s value of 0.42. Finally, the best extraversion combination metric for Físchlár-DT 2 (Extravert interface) was *Most Extravert*, with an r_s value of 0.22.

From statistical norms, we saw that the critical value for n degrees of freedom (n = sample size - in our case 18), and at $\alpha = 0.05$ for a two-tailed test ($\alpha = 0.025$ for a one-tailed test) is ± 0.476. Hence, none of the metrics used had a significant correlation to the performance of dyads in any of the systems. This finding supported the preliminary graphical findings in relation to this data. Both the Físchlár-DT 1 (Find 10 rule) and Accuracy Memory Game had values very close to the critical value, but were still deemed insignificant. Físchlár-DT 2 (Introvert interface) conveyed the weakest relationship between combined extraversion and performance. Table 5.1 summarises the results discussed above.

Since the correlations between these metrics were quite weak, our conclusion here is that other personality factors, not just *Extraversion*, may have had an impact on these dyads' performances. In fact, it may be that *Extraversion*, in conjunction with another personality trait, may have some correlation to the performance of a dyad. We decided to study all of the other personality traits in the next question, in order to see if correlations existed between dyads' performances on collaborative systems and participants' personality profiles.

	Best Combination Metric	r_s	**Statistically Significant?**
Accuracy Memory	*E-dist*	0.42	No
Speed Memory	*Most Introvert*	-0.31	No
Físchlár-DT 1 (10 min)	*Avg. Extraversion*	0.25	No
Físchlár-DT 1 (Find 10)	*Avg. Extraversion*	-0.44	No
Físchlár-DT 2 (Int)	*Most Introvert*	-0.20	No
Físchlár-DT 2 (Ext)	*Most Extravert*	0.22	No

Table 5.1: *Correlations of performance and extraversion*

We carried out the same type of analysis for our interaction data for all dyads and all systems - that being the touch-points recorded and the interaction instances that we annotated from the CCTV footage of the experiments. Again, we used the *Extraversion* combination metrics that we listed above and again, a preliminary bar-graphing of these results showed only random scatterings.

We also used the Spearman correlation coefficient to statistically determine whether significant relationships existed between the variables. Taking first the interaction instances, our variables were the *Extraversion* combination metric rankings and the rank of interaction instances for all dyads. We found that the *Average Extraversion* metric showed the strongest relationship for Físchlár-DT 1 (10 minute-rule) and Físchlár-DT 2 (Introvert), with r_s values of 0.51 and 0.70 respectively.

Most Extravert was the best indicator of a relationship in Speed Memory, Físchlár-DT 1 (Find 10 rule) and Físchlár-DT 2 (Extravert) with r_s values of 0.39, 0.34 and 0.74, while *Most Introvert* was the best metric for determining a relationship on the Accuracy Memory system with an r_s value 0.61.

Due to the fact that interaction data was corrupted for some dyads in some systems, our n values and subsequently our critical values were different for each system. This corruption was caused by a fault with the camera, which was not initially detected, since all video annotation of the camera footage was carried our after all of the first batch of user-experiments were completed. This fault was discovered in time for the second batch of experiments and repaired. For both Collaborative Memory systems, our sample size was n=17 (the video for Dyad 5 was corrupt). With an α of 0.05 (two-tailed test), our critical value here was 0.507.

For Físchlár-DT 1, our n was 17 (the video for Dyad 9 was corrupt), hence our critical value is 0.507. Videos for Dyads 1 and 13 were corrupt for the experiment using the Físchlár-DT 2 Extravert interface and so our n here was 16, with a critical value of 0.507. Lastly, our Físchlár-DT 2 Introvert system had an n of 15 (videos for Dyads 1, 11 and 13 were missing) and a critical value of 0.545.

It is clear from this that Físchlár-DT 2 (Introvert) had a strong significant positive correlation between the *Average Extraversion* metric and interaction instances, indicating that users with higher average *Extraversion* values have a greater number of interaction instances on this system interface. This means that dyads with a high level of average *Extraversion* communicated better on both these interfaces than those that had a lower average. Similarly Físchlár-DT 2 (Extravert) shows a highly significant correlation between *Most Extravert* and interaction instances, indicating that dyads with at least one highly extraverted member had more interaction instances. This would appear to support previous psychological research concerning people with high levels of *Extraversion* i.e. that they are talkative and sociable.

The Accuracy Memory System showed a significant relationship between the *Most Introvert* metric and a lower number of interaction instances. This indicated that those dyads with a member whose level of extraversion was very low, would have less communication instances with their partner, than those dyads with less introvert members - a logical finding.

	Best Combination Metric	**r_s**	**Statistically Significant?**
Accuracy Memory	*Most Introvert*	0.61	Yes
Speed Memory	*Most Extravert*	0.39	No
Físchlár-DT 1 (10 min)	*Avg. Extraversion*	0.51	Yes
Físchlár-DT 1 (Find 10)	*Most Extravert*	0.34	No
Físchlár-DT 2 (Int)	*Avg. Extraversion*	0.70	Yes
Físchlár-DT 2 (Ext)	*Most Extravert*	0.74	Yes

Table 5.2: *Correlations of interaction instances and extraversion*

The reason that the Físchlár-DT 1 (Find 10 rule) and Speed Memory Game show slightly less than significant correlations between any of our *Extraversion* combination schemes and interaction instances, could be because of the nature of the tasks. While communication was still necessary, the participants would have been in a greater hurry to complete the task and so may have been focusing more on their own situation rather than taking their time and interacting more with their partner. Also, since these are overall figures for interaction instances, and the length of time taken to complete the task varied, those who took longer to complete the task would be more likely have more interaction instances, regardless of personality type.

As a result of this, we decided to look at interactions on a per minute basis for the Accuracy Memory, Speed Memory and Físchlár-DT 1 (Find 10 rule), so that we could compare dyads in a normalised manner. Here, we found that the strongest dyad extraversion combination schemes were *Average Extraversion* for each of these systems, with r_s values of 0.65, 0.50 and 0.56 respectively. This shows that the greater the dyad's level of extraversion, the more likely they were to communicate either verbally or through gestures in the Accuracy Memory game and Físchlár-DT 1 (Find 10 rule) systems.

Our r_s value for the Speed Memory system was just outside the critical value of 0.507, hence we cannot statistically verify that the same outcome seen in Accuracy Memory and Físchlár-DT 1 (Find 10 rule) is likely to happen. However, due to the closeness of the values, we can suggest that this type of outcome is more likely to occur. The *Most Introvert* metric, which also showed a significant positive r_s value of 0.65 for the Accuracy Memory system, indicates that those dyads that contained at least one very introvert member were less likely to communicate than those with more extraverted members.

Hence, we can see that the *Extraversion* personality trait does seem to have a more significant overall impact on the communication of the group, something that we would expect given the social and collaborative nature of these systems. A summary of these results is provided in Table 5.2.

Next, we took the touch-point data recorded for each dyad on all of our collaborative systems. Here, our variables were the *Extraversion* combination metric rankings and the ranking of touch-point data. Once again, graphing by our devised combined extraversion metrics illustrated no visual correlations. Using our Spearman rank correlation method, we discovered that *E-Dist* presented the strongest correlations of the variables on 1 of the systems - our

Físchlár-DT 1 system. This gave us an r_s value of 0.40. Our *Most Introvert* metric proved to be the best combination metric in determining a relationship with user touch-points on the remainder of our systems, with r_s values of -0.48, -0.48, -0.35 and -0.22 calculated for the Accuracy Memory, Speed Memory game, Físchlár-DT 2 (Introvert) and Físchlár-DT 2 (Extravert) interfaces respectively.

Here our critical value for n=18, $\alpha = 0.05$ (two-tailed test) is 0.476. Therefore, only our Accuracy Memory and Speed Memory systems conveyed significant relationships between dyads with more introverted members, having a higher number of touch-points. Table 5.3 displays a summary of this information.

	Best Combination Metric	$\mathbf{r_s}$	**Statistically Significant?**
Accuracy Memory	*Most Introvert*	-0.48	Yes
Speed Memory	*Most Introvert*	-0.48	Yes
Físchlár-DT 1 (both)	*E-dist*	0.40	No
Físchlár-DT 2 (Int)	*Most Introvert*	-0.35	No
Físchlár-DT 2 (Ext)	*Most Introvert*	-0.22	No

Table 5.3: *Correlations of touch-points and extraversion*

Again, to counteract the fact that the time taken to complete the Accuracy Memory, Speed Memory, and Físchlár-DT 1 systems varied, we looked at the touch-points on a per minute basis in order to normalise our data. We discovered that the *Extraversion* combination scheme with the strongest relationships for Accuracy Memory, Speed Memory and Físchlár-DT 1 were *E-Dist*, *Average Extraversion* and *Most Extravert*, with r_s values of 0.43, 0.06 and -0.27 respectively. These values did not denote any significant correlations for any of these metrics with the number of touch-points per minute on these systems.

These findings have an impact on the design of multi-user tabletop interfaces, where dyads with levels of *Extraversion* that directly impact their interaction (as reflected in Tables 5.2 and 5.3) would require interfaces that support this interaction. This would impact the types of widgets used, the display and placement of objects on the interface.

While some of these r_s values conveyed strong relationships, particularly in the correlations of *Extraversion* and interaction instances, very few others denoted particularly strong correlations between the ranks. In fact, our extraversion combination metrics showed no correlations to performance in any of our collaborative systems. Hence, we decided to look at the remaining four of the "Big Five" personality traits to see if they provided us with stronger correlations to dyad performance on these systems.

Competitive Systems

For the competitive systems, Pop-A-Bubble and Competitive Memory Game, we attempted to identify a correlation between the most extraverted member of the dyad and the overall winners of the games. We looked at the results of the most extraverted members of the dyads to see if they won more games. Similarly, we also looked at the results of the most introverted members of the dyads.

The results of Competitive Memory Game showed that, in fact, the more introverted members of the dyads won more of the games than the more extraverted members on both the Introvert and Extravert interfaces. This majority was greater in the Introvert Interface with 67% of more introverted users winning outright and a further 11% scoring a draw with their more extraverted opponent. When using the Extravert interface, 44% of more introverted dyad members won games outright and 28% scored a draw. Hence, players that were more introverted than their opponent tended to perform better on this competitive system.

Our Pop-A-Bubble results showed once again, that the more introverted dyad members won more games than the more extraverted members in both the dual track-bar (60%) and the first single track-bar (i.e. where one user controlled the track-bar) games (73%). A higher percentage of the more extraverted members won the second single track-bar (i.e. where the other user controlled the track-bar) version of Pop-A-Bubble (73%). One participant, Dyad 8 User 1, with an average level of *Extraversion* (57% extravert) commented on his post-experiment questionnaire that "This brought out my competitive side, even though I'm not competitive. But it was fun!". This suggests one explanation as to why more introverted people performed so well at this game - they may have been forced to overcome their inhibitions, since the game was so fast-paced and competitive.

We looked also at the ages of participants to see if this was a factor in the success of participants. All participants fell into one of three groups – younger than 20 (13 participants), 20-24 (16 participants) and 25-29 (7 participants); the youngest participants being aged 17. Table 5.4 shows the breakdown of performance in terms of age of the dyads.

One can see from this table that the percentage of people winning and losing in the games are fairly even for the Pop-A-Bubble variations. The only notable difference arises in 25-29 year olds, where 67% win on the Extravert Memory game and only 33% lose. Also, slightly more 25-19 year olds win on the Introvert Memory than lose their games.

One of the reasons that this breakdown is relatively even, is because most dyad members were in the same age grouping as their partners. Only four dyads had members who ticked a different box for their age grouping from their partner. These were Dyads 1 (User 1: 20-24 and User 2: 25-29), 10 (User 1: 20-24 and User 2: Younger than 20), 14 (User 1: 25-29 and User 2: 20-24) and 17 (User 1: 20-24 and User 2: 25-29)). So for instance, if both members of a dyad are in the "Younger than 20" age-group and one member of a dyad wins, the other must lose and so the effect is balanced evenly.

Game	**<20 (Win)**	**<20 (Lose)**	**20-24 (Win)**	**20-24 (Lose)**	**25-29 (Win)**	**25-29 (Lose)**
Pop-A-Bubble Dual	43%	57%	50%	50%	57%	43%
Pop-A-Bubble S1	57%	43%	50%	50%	43%	57%
Pop-A-Bubble S2	43%	57%	54%	46%	43%	57%
Extravert Memory	50%	50%	46%	54%	67%	33%
Introvert Memory	54%	46%	43%	57%	60%	40%

Table 5.4: *Performances by Age*

Our competitive systems did not show a relationship between touch-points and participants with a higher level of extraversion than their opponent. In fact, those who had a higher level of extraversion than their opponents, were likely to have more touch-points than their opponents about 47% of the time on average (with a standard deviation of 5%). On the Competitive Memory system (both interfaces), more extravert players had more touch-points than their opponents, on average, 58% of the time (with a standard deviation of 3%).

In terms of interaction instances, we noted that more extraverted dyad members communicated more on the Extravert interface. 11 of the 18 more extraverted dyad members communicated more on the Extravert interface, in comparison to 7 on the introvert interface. However, this does not provide us with a significant majority - it merely suggests to us that the participants who are more extravert than their opponents, were more likely to communicate more than their opponents on the extravert interface. We did not look at communication on the Pop-A-Bubble system, since the games were particularly short (30 seconds each) and very little communication was noted in general.

While these results do not significantly relate to people with either low or high levels of *Extraversion*, they do give an indication that the *Extraversion* personality trait has more significance in performance on these competitive systems than in our collaborative systems.

In summary, while some of these r_s values conveyed strong relationships in our collaborative systems, particularly in the correlations of extraversion and interaction instances, very few others denoted particularly strong correlations between the ranks. In fact, our extraversion combination metrics showed no correlations to performance in any of these collaborative systems. Hence, we decided to look at the remaining four of the "Big Five" personality traits in response to the proceeding hypothesis questions to see if they could provide us with stronger correlations to dyad performance on these systems. We do this when answering the next hypothesis sub-question.

In conclusion, we should not focus solely on the Extraversion trait in attempting to correlate personality with dyad performance and interaction data. Extraversion was not significantly related to performance or normalised touchpoints in our collaborative systems, thought it did indicate that this trait was important for performance on competitive systems. Extraversion did significantly relate to interaction instances on many of our collaborative systems.

Q 2. Do the remaining "Big Five" personality traits affect the performances of dyads ?

To answer this question, we looked at each of the personality traits in turn, to determine whether they had an effect on the performances of the dyads on all of our systems. Performance here refers to the number of games won on Pop-A-Bubble, the number of mismatches and the length of time to complete the Collaborative Memory Game. In Competitive Memory, performance refers to the number of games won by each player. In Físchlár-DT 1 (10 minute rule) and both Físchlár-DT 2 interface variants, performance refers to the number of relevant shots found and for Físchlár-DT 1 (Find 10 rule), performance refers to the time taken to find 10 relevant shots.

We took a similar approach here as we did in determining if the *Extraversion* trait had a significant impact on performance. We examine firstly our collaborative systems, followed by our competitive systems.

Collaborative Systems

Similar to Question 1 above, we needed to combine the scores of both dyad members for each of the remaining four personality traits in some way, so that we could compare these to the overall performances of the dyads. In order to combine the Openness to Experience, Conscientiousness, Agreeableness and Neuroticism traits of the two members of each dyad, we looked at the absolute difference between the two members' scores on each of these traits. We also looked at the average of the two members' scores, the dyad member with the highest score and the dyad member with the lowest score along all these four traits. We then conducted a Spearman rank correlation between the ranks of each resulting trait combination and the performance ranks of the dyads on each of the systems.

Since our sample size here was 18, our critical value for $\alpha = 0.05$ (two-tailed test) was 0.476. Table 5.5 summarises the personality traits that we found to significantly correlate to performance on each of our systems. We note here that the *O-Dist*, *C-Dist*, *A-Dist* and *N-Dist* metrics listed are calculated in the same way as the *E-Dist* metric, except these refer to the differences in dyad members' scores along the *Openness to Experience*, *Conscientiousness*, *Agreeableness* and *Neuroticism* traits respectively.

	Related Traits	**r_s value**
Accuracy Memory	*Conscientiousness (Least Con.)* *Agreeableness (Least Agree., Avg. Agree.)*	-0.53 -0.55, -0.53
Speed Memory	—	—
Fís-DT 1 (10 min)	*Openness (O-Dist)*	-0.49
Fís-DT 1 (Find 10)	—	—
Fís-DT 2 (Int)	*Openness (Least Open)* *Conscientiousness (Least Con.)*	-0.54 -0.47
Fís-DT 2 (Ext)	*Openness (O-Dist), (Most Open)*	-0.64, -0.47

Table 5.5: *Traits significantly related to performance and associated combination metric for each system.*

From this table, we can see that none of the personality traits were related to performance on either of the 2 systems that required users to complete the task quickly. Recall that Speed Memory required users to complete the task as quickly as possible, while the Físchlár-DT 1 (Find 10) system, encouraged users to find 10 relevant shots as quickly as possible in order to complete the task. The discovery that none of the "Big Five" personality traits affected performance in these scenarios is very interesting.

In terms of the remaining search systems and interfaces – Físchlár-DT 1 (10 minute), Físchlár-DT 2 (Introvert interface) and Físchlár-DT 2 (Extravert interface), we can see that *Openness to Experience* is an important trait in relation to performance. In particular, the negative correlation of increasing *O-Dist* and increasing performance on Físchlár-DT 1 (10 minute) and Físchlár-DT 2 (Extravert interface) tells us that dyads, whose members have high *O-Dist* values perform better than dyads whose *O-Dist* values are small (i.e. are more similar).

We also see that there is a significant correlation between dyads containing at least one member who has a high level of *Openness to Experience* and low performance in Físchlár-DT 2 (Extravert interface). Físchlár-DT 2 (Introvert interface) displays a negative correlation between dyads containing members with low *Openness to Experience* scores and increasing performance i.e. dyads containing members with low *Openness to Experience* scores performed better than those with higher scores along the same trait.

The negative correlation between increasing performance ranks and increasing *Least Conscientiousness* ranks (i.e. rank of dyad members with a lower *Conscientiousness* score than their partners) for Físchlár-DT 2 (Introvert interface) is also a surprising finding, since we believed that this system required both users to think before they acted. Since the functions were duplicated on this interface, there was more potential for users to interrupt and intercept their partner's work, by playing over their videos, covering over their search results, etc. In the other Físchlár-DT system interfaces, widgets were shared and so people did not encounter the need to think so much about their actions – they were more aware of each other and could see if their partner was using a particular function. Hence, we expected that having a less impulsive personality would have been important in this system. Once again though, this did not appear to be the case here.

Both *Conscientiousness* and *Agreeableness* were important in the Accuracy Memory system. Here, decreasing performance was negatively correlated to both *Conscientiousness (Least Conscientious)* and *Agreeableness (Least Agreeable and Avg. Agreeableness)*. We recall that low figures for performance on Accuracy Memory indicate better performance (i.e. fewer mismatches). Hence, these resulting correlations were much more intuitive, since it would be important that dyad members were both conscientious (i.e. they lacked impulsiveness and thought before they acted) and agreeable, so that they made the fewest errors possible.

We can also see that *Neuroticism* had no effect on performance in any of our collaborative systems – the same result as that found for the *Extraversion* trait in Question 1. Initially, we had thought that *Extraversion* would have been the personality trait to most significantly impact the performances of dyads. However, we see from Table 5.5 that other traits have far greater significance in affecting performance on our collaborative tasks, while *Extraversion* has no statistically significant impact on performance. We now examine the four personality traits examined here in relation to dyad performance on our competitive systems.

Competitive Systems

Analysing personality traits in relation to performance on competitive systems involved identifying whether the overall winners in these systems generally had either higher or lower scores along those personality traits.

We commence by looking at the winners of our Pop-A-Bubble system variants. On the Dual Track-bar system, 60% of the winning participants were more Open to Experience than their opponents, 53% were more Conscientious, just 33% were more Agreeable and 53% were more Neurotic. 7% scored a draw on this interface. On the first Single Track-bar system, 26% of winners were more Open to Experience, 60% were more Conscientious, 67% were more Agreeable and again, 53% were more Neurotic (both members of one dyad had the same score along the *Neuroticism* trait).

Finally, on the second Single Track-bar system, 40% of winners were more Open to Experience, 73% were more Conscientious, 53% were more Agreeable and again, 40% were more Neurotic. On average, the only personality trait that appeared to have a greatest effect on winning was *Conscientiousness*, with an average of 62% across all variants – almost twice as much as the average number of participants who were less *Conscientious* than their opponent (36%). This is surprising, since higher levels of *Conscientiousness* are associated with a lack of impulsiveness, which is the opposite to what winning in this system requires.

Next, we look at our Competitive Memory systems – The Introvert Memory Game and Extravert Memory Game. We discovered in both these interfaces, that winners were fairly evenly distributed between higher and lower scores along each trait. For our Introvert Memory game, 33% were more Open to Experience than their opponents, while 39% were less Open to Experience. 39% were more Conscientious while 33% were less, 39% were more Agreeable, while 33% were less and finally, 39% were more Neurotic than their opponents, while 33% were less. 28% of participants scored a draw on this interface.

For our Extravert Memory game, 50% were more Open to Experience than their opponents, while 39% were less Open to Experience. 56% were more Conscientious while 33% were less, 33% were more Agreeable, while 56% were less Agreeable and 39% were more Neurotic than their opponents, while 50% were less. 11% of participants scored a draw on this interface. Here, we can see that a much greater proportion of winners were more *Conscientious* than their opponents, in comparison to the other traits – a similar finding to that of the Pop-A-Bubble system.

Hence, we can see from the above that *Conscientiousness* appears to have the greatest effect on winning in two of our competitive systems i.e. Pop-A-Bubble and Extravert Memory game. No large differences were found in the remaining personality traits of winners on any of these systems. We also recall from Question 1 above, that low levels on *Extraversion* also had an effect on winning on our competitive systems.

These results in terms of the performance data as related to personality traits provide a meaningful contribution to the area of Personality Psychology. The inferences drawn indicate that personality has an important role to play in predicting the performances of dyads on the tasks we used that didn't impose a timing constraint.

In conclusion, combinations of four of the "Big Five" personality traits are important in determining performance on the collaborative systems that do not impose timing constraints. *Agreeableness* and *Conscientiousness* are important in Accuracy Memory, while *Openness to Experience* is important on Físchlár-DT 1 (10 Minute) and both Físchlár-DT 2 interfaces. *Conscientiousness* is important on Físchlár-DT 2 (Introvert interface). None of the personality traits correlate to Speed Memory or Físchlár-DT 1 (Find 10). *Conscientiousness* and *Extraversion* are important traits in relation to performance on our competitive systems.

Q 3. Do dyads that are more similar in terms of their personality composition outperform dyads containing very different personality types ?

In order to look at this question, we used a personality "congruence" score (Shell and Duncan, 2000). Prior to this work, we had conducted a study which attempted to correlate the personality match of dyads (using the MBTI method of profiling personality) to performance (Smeaton et al., 2007). Here we attempted to do the same thing, but rather than using personality match based on MBTI, we attempt to correlate"The Big Five" personality questionnaire (Johnson, 2008) results obtained from a much larger sample of dyads, with their performance. In calculating an overall score for personality congruence, we added together the absolute differences in each trait, that each dyad member scored from their partner i.e. the *E-Dist*, *O-Dist*, *A-Dist*, *C-Dist* and *N-Dist* that we used in Questions 1 and 2 earlier. By adding these resulting values together, we could obtain a measure of how similar the dyad members were – the smaller the personality congruence value was, the more similar the dyad members were to each other in terms of their personality.

We used a Spearman rank correlation to determine whether the personality congruence scores of the dyads were correlated to the performance of the dyads on each of our systems. If we were to follow the law of similarity attraction, then we would expect that dyads, who were more similar in terms of their personality, would perform better than those who were more diverse. We do not examine our competitive systems, since there is no sense of "dyad performance" on a competitive system – performance is based on the individual's output or score.

Once again, our sample size here was 18, and at $\alpha = 0.05$ (two-tailed), our critical value was 0.476. Table 5.6 displays the r_s values calculated for ranks of performance and personality congruence on each of our collaborative systems and their interface or rule variants.

Form this table, we can see that only the Accuracy Memory game shows

	r_s value	Statistically Significant?
Accuracy Memory	0.52	Yes
Speed Memory	0.36	No
Físchlár-DT 1 (10 min)	-0.05	No
Físchlár-DT 1 (Find 10)	0.43	No
Físchlár-DT 2 (Int)	-0.20	No
Físchlár-DT 2 (Ext)	-0.06	No

Table 5.6: *Correlations of personality congruence and performance*

a statistically significant correlation between performance and personality congruence. None of the other systems show a significant correlation, although the Físchlár-DT 1 (Find 10) system has a value extremely close to the critical value.

Conclusion: Since personality congruence was only related to dyad performance in one system (Accuracy Memory game), we cannot say that, generally, dyads with similar personality types outperform dyads with differing personality types.

Q 4. Is the interaction recorded among dyads related to their personality composition ?

We use the same approach to answering this question as we did when answering Question 2 earlier. However, here we look at the Openness to Experience, Conscientiousness, Agreeableness and Neuroticism combination metrics in relation to the number of interaction instances and touch-points of the dyads. Once again, we investigate interaction instances with regard to our collaborative systems only, since our competitive systems had very few interaction instances. We do, however, examine the touch-points of the dyads in relation to personality on our competitive systems.

Taking firstly the interaction instances of the dyads, we conducted a Spearman rank correlation of the ranks of dyad interaction instances and the dyad ranks along trait combination schemes (see Question 2 for more details on these schemes). Once again, our sample sizes (n) for interaction instances are different for some of the systems. Here, n = 17 for Accuracy and Speed Memory, as well as both Físchlár-DT 1 rules, n = 15 for Físchlár-DT 2 (Introvert interface) and n = 16 for Físchlár-DT 2 (Extravert interface) systems. Hence, our critical values at $\alpha = 0.05$ (two-tailed test) are 0.507 for Accuracy Memory, Speed Memory, Físchlár-DT 1 (both rules) and Físchlár-DT 2 (Extravert), and 0.545 for Físchlár-DT 2 (Introvert).

Table 5.7 summarises the significant correlations found between each personality trait and the normalised number of interaction instances incurred by the dyads. We see from this table that only the Físchlár-DT 2 interfaces showed significant correlations.

Físchlár-DT 2 (Introvert interface) displays a significant negative correlation between dyads containing members scoring lower on the *Neuroticism* personality trait and increasing number of interaction instances i.e. dyads where both members were highly neurotic had fewer interaction instances. Físchlár-DT 2 (Extravert interface) showed a significant, positive correlation between in-

	Related Traits	r_s value
Accuracy Memory	—	—
Speed Memory	—	—
Fís-DT 1 (10 min)	—	—
Fís-DT 1 (Find 10)	—	—
Fís-DT 2 (Int)	*Least Neurotic*	-0.59
Fís-DT 2 (Ext)	*Avg. Agreeableness*	0.55

Table 5.7: *Traits significantly related to interaction instances and associated combination metric for each system.*

creasing *Average Agreeableness* and increasing interaction instances i.e. dyads whose members were on average more agreeable, had more interaction instances. These were intuitive findings.

We also saw from Table 5.2 (see Question 1 above), that *Extraversion* was significantly correlated to four out of the six collaborative systems that we used in our study. Hence, *Extraversion* is the most important trait with regards to interaction instances.

Finally, in determining whether personality is related to the interaction of the dyads, we look at the touch-points recorded for each dyad on each of the systems in Speed Memory. Table 5.8 below displays the significant correlations between the respective personality traits and the touch-points of the dyads.

Here we see that *Conscientiousness* is significantly related to the number of touch-points for both Collaborative Memory game rules. However, the positive relationship for Accuracy Memory indicates that dyad members, whose level of *Conscientiousness* was similar (i.e. a low *C-Dist*), had fewer touch-points per minute than those who had very different levels of *Conscientiousness* (i.e. a high *C-Dist*).

However, an increasing level in the number of touch-points for each dyad was negatively correlated to both *Average Conscientiousness* and *Least Conscientious.* This implied that dyads who had a high average Conscientiousness or where both members had high levels of *Conscientiousness* had fewer touch-points than those with a lower average or at least one member with a very low score on the *Conscientiousness* trait. This was a logical finding, since highly conscientious people are known to show a lack of impulsiveness and are generally more thoughtful. This lack of impulsiveness is reflected by the fewer touch-points.

We also see that *Neuroticism* is significantly correlated to the number of touch-points, where dyads who have similar levels of *Neuroticism* have fewer touch-points than those whose members have very different levels of *Neuroticism*. In addition, dyads containing at least one member with a high level of *Neuroticism* had significantly more touch-points than those dyads whose members had lower levels of *Neuroticism*.

We can see that *Neuroticism* has a significant negative correlation to increasing number of dyad touch-points on Físchlár-DT 1 along the *N-Dist* metric i.e. dyads whose members' level on *Neuroticism* was similar had more touch-points than those whose members' level of *Neuroticism* was very different. *Openness to Experience* was negatively correlated to increasing number of touch-points on Físchlár-DT 2 (Introvert Interface), along the *Most Open* metric, meaning that dyads who had at least one dyad member with a relatively high level of openness had fewer touch-points than dyads where both members had relatively low openness scores.

Finally, *Agreeableness* displayed a significant inverse correlation to increasing number of touch-points on the Físchlár-DT 2 (Extravert interface), along both the *Most Agreeable* and *A-Dist* metrics. This implies that dyads with at least one relatively more agreeable member had fewer touch-points, while dyads whose members had more similar levels of *Agreeableness* had more touch-points than dyad members who had very different levels of *Agreeableness*.

We look now at our **competitive systems**. Table 5.9 provides us with a summary of players who had a greater number of touch-points on each of the systems and percentages of these who were more or less open to experience,

	Related Traits	**r_s value**
Accuracy Memory	*Conscientiousness (C-Dist)*	0.49
Speed Memory	Conscientiousness (Avg), (Least)	-0.62, -0.48
	Neurotic (N-Dist, Most Neurotic)	0.53, 0.5
Fís-DT 1	*Neuroticism (N-Dist)*	-0.59
Fís-DT 2 (Int)	*Openness (Most Open)*	-0.49
Fís-DT 2 (Ext)	*Agreeable (Most), (A-Dist)*	-0.51, -0.48

Table 5.8: *Traits significantly related to touch-points for each system and associated combination metric.*

	Pop (Dual)	**Pop (Single 1)**	**Pop (Single 2)**	**Introvert Memory**	**Extravert Memory**
More Open	40%	53%	47%	39%	39%
Less Open	60%	47%	40%	50%	61%
More Conscientious	60%	47%	67%	39%	44%
Less Conscientious	40%	53%	20%	50%	56%
More Agreeable	53%	67%	40%	39%	67%
Less Agreeable	47%	33%	47%	50%	33%
More Neurotic	33%	53%	13%	44%	44%
Less Neurotic	60%	40%	67%	44%	50%

Table 5.9: *Influence of personality traits on touch-points in each system.*

conscientious, agreeable and neurotic than their opponent.

We can see from this table, that the only traits that showed the greatest majorities in terms of trends in participants with more touch-points were *Neuroticism* on the Pop-A-Bubble Dual system, *Agreeableness* on the first Single Track-bar Pop-A-Bubble, *Neuroticism* on the second Single Track-bar Pop-A-Bubble and *Agreeableness* on Extravert Memory. Firstly, players with more touch-points on the Pop-A-Bubble Dual system were almost twice as likely to be less neurotic than their opponent. On the first Single Track-bar Pop-A-Bubble, players with more touch-points were just over twice as likely to be more agreeable than their opponent. Participants with more touch-points were more than three times more likely to be more conscientious on the second Single Track-bar Pop-A-Bubble. Finally, players with more touch-points on the Extravert Memory interface were more than twice as likely to be more agreeable than their opponent. We also recall from Question 1 above, that *Extraversion* had minimal connection to number of touch-points on these systems.

Hence, we can see that personality had a significant relationship to the interaction of the dyads in terms of their communication (interaction instances) and their physical interaction with the tabletop (i.e. their touch-points) on the collaborative systems. With regards to interaction instances, *Extraversion* was the personality trait that had a major impact, while the other traits were not effective any of the interfaces but the Físchlár-DT 2 system. However, four of the "Big Five" personality traits had a significant relationship to the number of dyad touch-points on at least one of the collaborative systems that we studied. Only *Agreeableness* and *Neuroticism* had an impact on touch-points in some of our competitive systems.

As in Q.1, these findings have an impact on the design of multi-user tabletop interfaces, where the combinations of dyad personality trait scores reflect those that significantly impact the interaction of dyads (as reflected in Tables 5.7, 5.8 and 5.9), would require interface designs that support this and take this into consideration. Again, this would impact the types of widgets used, the display and placements of objects on the interface.

Conclusion: Different personality traits are important for determining the level of interaction of dyads, both on our competitive and our collaborative systems. For our collaborative systems, *Neuroticism* was found to statistically affect interaction on Físchlár-DT 2 (Introvert Interface), while *Agreeableness* affected interaction on Físchlár-DT 2 (Extravert Interface). *Conscientiousness* was found to be significantly correlated to touch-points on both Collaborative Memory variations. *Neuroticism* was also statistically correlated to touch-points on Speed Memory and Físchlár-DT 1. *Agreeableness* had a significant impact on touch-points on Físchlár-DT 2 (Extravert Interface) and *Openness to Experience* was important on Físchlár-DT 2 (Introvert Interface). *Agreeableness* had a significant impact on touch-points on the first and Extravert Memory, while *Neuroticism* had an impact on touch-points on Pop-A-Bubble Dual. Hence, dyads with certain combinations of personalities interact more than others, though these combinations vary with each system.

Hypothesis Support

From looking at the answers to each of the sub-questions examined so far, we can now state that in our study, dyads exhibiting levels of *Conscientiousness*, *Agreeableness* and *Openness to Experience* performed significantly better on our collaborative systems, while *Conscientiousness* and *Extraversion* were more important traits in the winners of our competitive tasks. Hence, dyads with certain personality types perform better in tasks compared to others.

5.4 Hypothesis 2

Dyads with certain personality types will prefer and work better on certain interfaces

We look now to four hypothesis sub-questions in order to prove or disprove this.

Q 5. Do individuals within dyads develop a similar impression of a system ?

To answer this question, we look to our Post-Experiment questionnaire data for each of our systems, in order to see if dyad members reported the same preferences or if they passed similar comments and remarks on the system (questionnaire responses from one of our systems are supplied in Appendix A). We look at both direct questions posed, as well as opinion-scale ratings to statements we asserted about the systems. The scale contained seven points - Strongly Agree, Quite Agree, Little Agree, Neutral, Little Disagree, Quite Disagree and Strongly Disagree. All participants were requested to rate these statements in response to their experience in using the systems.

We take each of the systems in turn to see the variability that existed in the opinions of each of the participants within dyads. Firstly, we look at **Pop-A-Bubble**. When asked what their overall impression of the system was, the majority of participants reported that they liked the game. Some participants even used the same terms to describe their experience of using the system e.g. "fun", "very good", "entertaining" and "exciting". Both members of one dyad reported that they found it "confusing". One participant said that the system was "alright", while his partner stated that the system was "very good". In general, however, all other dyads reported a liking of the system.

When asked which version of the system they preferred, members within eight dyads agreed in their interface variant preference. Members within seven of these preferred the *Dual Score* version, with members of the remaining dyad preferring the version where they had control of the track-bar. One participant failed to respond and his partner reported that he noticed no difference in the interfaces. Another participant reported no preference and members within the remaining eight dyads differed in their preference.

The post-experiment questionnaire also listed five opinion-scale type statements. On observation of the dyads' ratings of these statements, we noted that all users gave very similar responses to the first statement ("The system is easy to use"). All except two participants agreed with this statement (Dyad 12 User 2 and Dyad 14 User 1 selected Neutral ratings). This reflects a comment made by Dyad 12 User 2, where he said "Relatively easy to use, although often failed to pickup what I was doing.".

Similarly, the statement "Learning how to use the system was easy" elicited mostly Strongly Agree or Quite Agree. The third statement, "The system response time was fast enough", elicited differing opinions within Dyads 4, 9, 11, 12 and 18. Participants within all of these dyads reported that they noticed that the system did not acknowledge when they touched the interface or touched a bubble e.g. Dyad 18 User 1 reported "I am better when the score bar is displayed on my side! Some bubbles don't pop". Hence the reason for the lower ratings. Some participants even reported a difference in response time between the dual track-bar and single track-bar, even though this was the same for both, showing that even simple differences in a system interface can give users a different impression of a system.

Statement 4, "The system interface allowed me to do the task efficiently" received mainly positive responses, although five participants from Dyads 2, 3, 11, 12 and 15 gave "Neutral" or "Little Disagree" responses. Again, upon reading earlier comments made by these participants, we found that they encountered difficulties in the system picking up their touch-points, which is probably why these ratings were given here. The final statement "It was easy to be aware of what the other person was doing" received more negative responses, with five participants either selecting the "Little Disagree" or "Strongly Disagree" rating. Five participants selected the "Neutral" rating. This is most likely because the game is particularly fast-paced, so participants had to concentrate on their own performance more. Also, the two Single-bar interfaces, as the name might suggest, had only one score/track-bar, hence those not sitting at the side where the track-bar was would be unaware of the scores of the other person.

Next we look at the responses to the post-experiment questionnaires of the **Collaborative Memory Game**. Once again, almost all participants reported that they liked the task they had been asked to complete. Some reported that they liked the novel, collaborative aspect of the game, since this has traditionally been played as a competitive game. An example of this was Dyad 8, User 2's comment - "Very entertaining :) Added a new 'light' on what I might have considered a solo playing game".

Users were then asked to rate their opinion of both the Accuracy and Speed task constraints. Out of the 18 dyads, members within 11 dyads expressed a similar opinion of the versions i.e. they preferred the same versions or expressed a liking of both versions. Dyad 2 User 1 had no opinion on the different rules (he rated them both as Neutral) and the remaining members within 6 dyads preferred differing versions. The only participant to express an extreme preference of one version over the other was Dyad 12 User 2, who quite agreed that he liked the Accuracy version of the system and quite disagreed with the statement that he liked the Speed version. We can see why he gave this rating from a comment made later in the questionnaire regarding coordination "We talked to set up a system and so to spread the things that have to be remembered. He was stupidly slow though". Hence the reason why he disliked the Speed version.

The statements that required opinion-scale ratings were examined next. The first of these was "I felt comfortable with the rules of the system", which the majority of users agreed with. One participant (Dyad 2 User 1) gave the Accuracy game a "Quite Agree" and the Speed version a "Neutral" rating, even though he expressed no preference in his liking of either system in the previous question. Dyad 12 User 2 gave the Speed version a “Little Disagree” rating, which was in keeping with his response as to whether he liked the systems.

The next statement “Learning how to use the system was easy” elicited mainly positive responses. This was in keeping with comments made by users when describing their overall impression of the system, where words such as “Intuitive Interface” (Dyad 2 User 2), “Easy to operate” (Dyad 9 User 1) were used. “The system interface allowed me to do the task efficiently” received a slightly more varied response, with three “Neutral” ratings, one “Little Disagree” rating and two “Quite Disagree” ratings to this statement for both versions. Members within the remaining 14 dyads had a similar opinion on this statement (i.e. they liked it).

“It was easy to be aware of what the other person was doing” received much more dissimilarity in the opinions of participants and their partners (members within seven dyads differed in their opinions). More negative ratings were supplied for the Speed version, which was unsurprising as we would have assumed that people, when placed under time pressure, would be more concerned about their own situation than observing the work of their partner.

The statements “I liked the layout of the game/interface” elicited mainly positive responses, with only one user giving a negative response to these (Dyad 12 User 2 gave a “Little Disagree” rating for this). “I found myself improving as I used the system” elicited more “Neutral” and negative responses than previous statements with members of Dyads 1, 3, 5 and 11 differing in their opinions, indicating that users were confident working on this system from the moment the task began.

We note that in general, users appeared to prefer working under the Accuracy rule than under the Speed rule for this system. There are a number of reasons why this was the case. For instance, some people like working at their own pace and executing tasks carefully, others like the challenge and “buzz” to be gained from working to a time deadline. From a personality perspective, *Conscientiousness* is the personality trait associated with impulsiveness (Johnson, 2008). Looking at the percentages of personality traits (see Appendix C), we can see that 18 participants scored above 50% on this trait and 18 participants scored below 50%. This implies that there is no obvious relationship between scores on this personality trait and interface preference.

We decided to look at the statistical relationship between user's ratings if each interface and scores on the *Conscientiousness* personality trait, using Spearman's rank correlation. The correlation was much stronger on for the Speed Memory interface (t-stat of -1.7) in comparison to Accuracy Memory (t-stat 0.36). Since our critical value here (for a two-tailed test, at a confidence of 0.05) is 2.042, both values are statistically insignificant. However, the negative correlation coefficient for Speed Memory does approach the critical value, indicating that people with higher levels of *Conscientiousness* (i.e. lower impulsiveness) rated this interface gave this a 1, 2 or 3 rating. While this is not statistically significant, it is still a surprising trend, since we would have assumed that impulsive people would generally have disliked this interface.

The **Competitive Memory Game** saw members within ten dyads agreeing on their preference of interface, with both members of five of these dyads preferring the extravert interface (Dyads 2, 8, 11, 15 and 16) and both members of the other five preferring the introvert interface (Dyads 4, 5, 10, 14, 17). Two participants from Dyads 1 and 3 failed to pick a preference in the systems (they expressed an equal liking for both) and participants in the remaining 6 dyads differed in their preferences. We did not name the systems "introvert interface" and "extravert interface" on our questionnaires, but rather called them *With Hints Button* and *With Automatic Hints*. This was so that users would not feel that certain aspects of their personalities were being tested directly, knowledge of which could have affected their preferences or performance.

In general, most users commented that it was easy to use, which is affirmed by their ratings of the first opinion-scale statement, though many also complained about the slowness of the system. "Learning how to use the system was easy" elicited a dissimilar response within Dyad 6. User 1 from this dyad disagreed with this statement for the *Introvert Interface*. "The system response time was fast enough" yielded varying within-dyad opinions for Dyads 5, 9 and 12. Just two dyads (Dyads 6 and 12) differed in their opinion on "The system interface allowed me to do the task efficiently".

Dyads 2 and 6 provided dissimilar within-dyad ratings for "It was easy to be aware of what the other person was doing". Dyad 2 User 1 gave the *Extravert Interface* a "Little Disagree" rating, while Dyad 6 User 1 gave the same rating to the *Introvert Interface*. Members within six dyads (Dyads 3, 5, 8, 11, 13, 15) disagreed in their responses to the statement "I liked the layout of the game/interface. (Colours used, etc)", with most varying opinions being on the *Extravert Interface.* This aspect of the interfaces is discussed further in Question 6 below.

"I liked competing against the other person" saw members within just four dyads varying in their ratings to this statement (Dyads 1, 4, 11 and 17), while members of Dyads 11 and 13 expressed differing opinions of the statement "I felt comfortable using the system". The final statement "I found myself improving as I used the system" elicited differing opinions within seven dyads (Dyads 1, 3, 4, 6, 11, 12 and 13).

We look also to the backgrounds of the users in terms of their age, gender and the frequency with which the participants play competitive games on a PC to determine if there are any visible trends in system opinion along these categories. We see from Figures 5.4, 5.5 and 5.6 that only participants who were part of the "younger than 20" category expressed a dislike for the Introvert Memory System.

Only participants who were aged between 20 and 24 expressed a dislike for the Extravert Memory System. Neutral ratings were relatively balanced across age groupings for both interfaces. Only people who played computer games

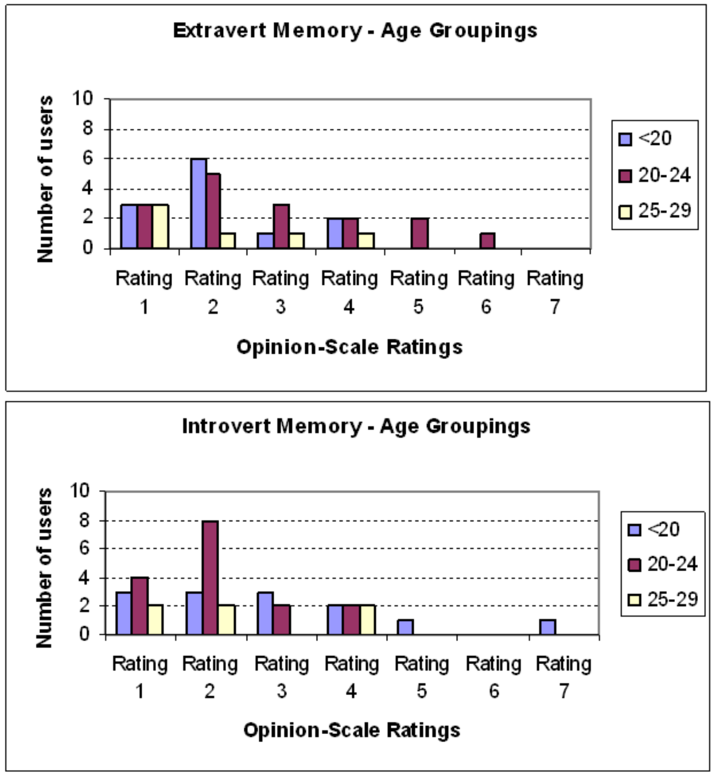


Figure 5.4: *Age and Opinion Rating for Competitive Memory*

seldom or once a week gave a dislike rating for both interfaces. In terms of gender, no females disliked the Introvert Memory interface, while one disliked the Extravert interface. Two males expressed a dislike of both the Extravert and the Introvert interface. However, as we see there are a far greater majority of males in the experimental sample set (33 males and 3 females); hence accurate conclusions cannot be drawn from this.

In **Físchlár-DT 1**, members within ten out of the eighteen dyads expressed a similar impression of the system (eight dyads liked it and two dyads disliked it). Only one member of Dyad 14 responded and members within the remaining seven dyads differed in their opinions of the system. Within Dyads 2, 3, 10, 12, 14 and 15, participants differed in their opinions on the opinion-scale type statement “The system is easy to use”. All but Dyad 13 agreed on the statement “Learning how to use the system was easy”.

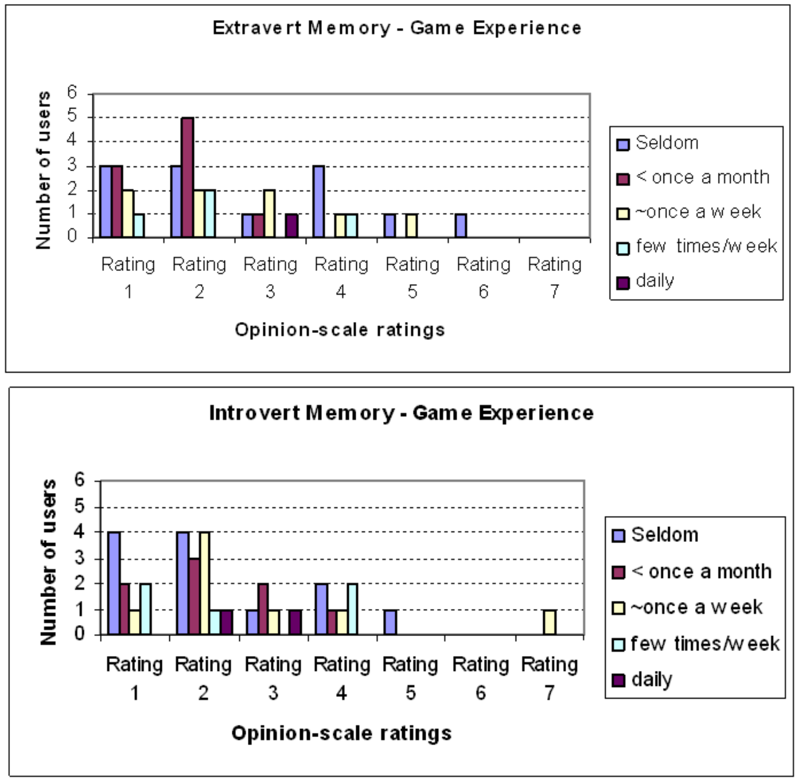
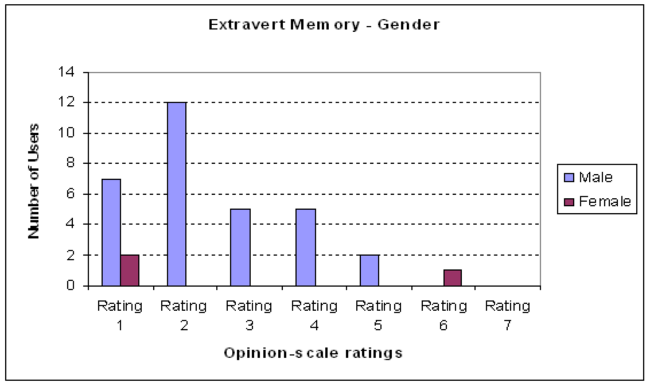


Figure 5.5: *Game Experience and Opinion Rating for Competitive Memory*

"The system response time was fast enough" elicited much less similarity in within-dyad responses. Just eight out of eighteen provided a similar within-dyad answer i.e. either both liked or disliked. All other dyads (Dyads 2, 5, 9, 10, 12, 13, 14, 16, 17 and 18) had differing within-dyad opinions. Statements about aspects of the interface that people disliked supported this opinion e.g. Dyad 2 User 1 simply stated "System response time was too slow.", while Dyad 4 User 1 offered a more in-depth explanation – "Slowdown when 2 people ran things. Slight lagging response time". Some of these were caused by technical issues that were fundamentally problems with the underlying DiamondSpin SDK and could not be solved by us.

Members within nine dyads had similar opinions on the statement "The system interface allowed me to do the task efficiently" (with five agreeing with the statement). Just eight dyads had similar within-dyad opinions on the statement



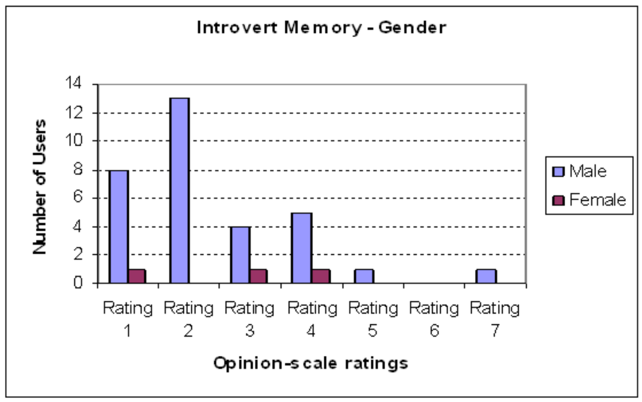


Figure 5.6: *Gender and Opinion Rating for Competitive Memory*

"I liked the layout of buttons and widgets", while members within twelve dyads shared a similar opinion on the statement "I liked the colours used". Fourteen dyads had similar within-dyad responses to the statement "It was easy to be aware of what the other person was doing".

We noted in our analysis of these opinion-scale responses, that in fact Dyad 13, User 2 gave mainly negative or "Neutral" responses to all statements. This was unsurprising however, as he expressed a rather extreme dislike of the system ("Didn't like this system at all and can't see anyone wanting to use it or enjoy it").

In **Físchlár-DT 2**, our questions did not use the names *Introvert Interface* and *Extravert Interface*, but rather we called the two interfaces *Separate Functions* and *Shared Functions* respectively. Both members of eleven dyads preferred the same interface, with members within nine of those dyads preferring the Separate Functions interface. User 1 from Dyad 18 liked both interfaces equally. Members from Dyads 2, 7, 8, 12, 13, 14, and 16 differed in their preference. Those that preferred the Extravert Interface preferred the fact that the interface was neater and had fewer hot-spots. For example, Dyad 2 User 2 stated that "The first one is very user-friendly and easy to use. The second one is a bit messy, too many options". Those who chose the Introvert interface as their preference liked the fact that they had their own set of hot-spots to work with e.g. Dyad 1 User 1 stated he preferred the Introvert Interface "Because I felt better with my own functions. It allowed me to concentrate on my own search if I wanted". Again, the mainly negative opinion scale ratings were concerned with the system glitches in the DiamondSpin SDK and not the interfaces themselves.

Fifteen dyads shared the same within-dyad opinion in their responses to the statement "The system is easy to use" (members within Dyads 3, 12 and 14 differed in their opinion), while members within sixteen dyads agreed in their opinion on the statement "Learning how to use the system was easy" (Dyads 3 and 12 differed). The remaining five statements "The system response time was fast enough", "The system interface allowed me to do the task efficiently", "I liked the layout of buttons and widgets", "I liked the colours used" and "It was easy to be aware of what the other person was doing" yielded many more differing opinions within dyads.

Table 5.10 summarises the overall agreement in interface preferences among dyads. Examining the dyads more closely we see that members within Dyad 10 share similar interface preferences and opinions on all five systems. Members of Dyads 1, 4, 11, and 15 agree on four of the systems and members of Dyads 3, 5, 6, 8, 16 and 17 agree on three out of the five. Dyads 2, 9, 13 and 18 agree on interface preferences for two systems, with members within Dyad 2 agreeing only on competitive interfaces and members within Dyads 9 and 18 agreeing only on collaborative interfaces. Dyads 7, 14 and 18 agreed on just

Systems	Dyads with same preferences
Pop-A-Bubble	Dyads 1, 2, –, 4, –, 6, –, –, –, 10, 11, —, 13, —, 15 —, —, —
Collaborative Memory	Dyads 1, –, 3, –, –, 6, 7, 8, 9, 10, 11, —, 13, —, 15, 16, —, —
Competitive Memory	Dyads –, 2, –, 4, 5, –, –, 8, –, 10, 11, —, —, 14, 15, 16, 17, —
Físchlár-DT 1	Dyads 1, –, 3, 4, 5, –, –, 8, –, 10, —, —, —, —, —, 16, 17, 18
Físchlár-DT 2	Dyads 1, –, 3, 4, 5, 6, –, –, 9, 10, 11, —, —, —, 15, —, 17, 18

Table 5.10: *Dyads sharing the same interface preference*

one interface, with both 7 and 18 agreeing on a collaborative interface. Dyad 12 failed to agree on any of their interface preferences.

As we can see from the above, members within our participating dyads agreed with each other more than half of the time, with the exception of Pop-A-Bubble. We had expected that the individuals with dyads would develop a similar impression of our collaborative system interfaces, possibly through comments that each might pass on the system, for example, if one person expresses that he/she likes a certain widget and finds it useful, then this might cause their partner to be drawn to this and also find it useful. We also thought that individuals within dyads would generally have different impressions of the competitive system interfaces, since they worked on an individual basis in the competitive tasks. However, there was very little difference between agreement on the competitive system interfaces and agreement on collaborative system interfaces. In fact, the average number of dyads that had within-dyad agreement on their interface preferences for the competitive systems (Pop-A-Bubble and Competitive Memory) was nine, while on the collaborative system interfaces, the average was ten.

Conclusion: Since dyads agree with each other in terms of their interface preferences just over half the time, we cannot confidently say that individuals within dyads generally develop a similar impression of a system interface.

Q 6. Do users prefer interfaces which model their personality along the *Extraversion* trait ?

As stated in Chapter 4, we gave our Competitive Memory Game and Físchlár-DT 2 system variants a "personality" i.e. one system interface variant portrayed *extravert* characteristics, while the other portrayed *introvert* characteristics, based on previous research by Reeves and Nass, 1996. Hence, to answer this question, we look at our Competitive Memory Game and Físchlár-DT 2 system variations, to see if users who were more *extravert* or more *introvert* in personality, liked interfaces portraying a similar trait.

Firstly, we looked at the post-experiment questionnaire responses supplied by dyads for the Competitive Memory system. We discovered that 12 out of 36 participants preferred the layout and colours on the Introvert Interface (six of these participants were from the same dyad), while two participants preferred the layout and colours on the Extravert interface. The remaining participants liked the layout and colours equally, with the exception of Dyad 11 User 2, who disliked them.

In order to examine this more closely, we looked at the actual opinion-scale ratings that each user gave for each interface, as well as their opinion-scale ratings for the colours and layout of each interface variant - very important components of the "personality" of the interfaces. We conducted a Spearman rank correlation between the users' opinion-scale ratings of the interfaces and their own scores on the *Extraversion* personality trait. We anticipated a positive correlation between increasing extraversion and increasing opinion-scale ratings on the Introvert interface (note: a rating of 1 represents "Strongly Agree", while a rating of 7 represents "Strongly Disagree"). For the Extravert interface, we believed that there would be a negative correlation between increasing *Extraversion* and increasing opinion-scale ratings, since increasing opinion-scale ratings indicate greater dislike of the system i.e. 1 = Strongly Agree and 7 = Strongly Disagree.

Our resulting r_s values on the Introvert interface were -0.17 for the ranks of users' ratings of their opinion of the interface and their *Extraversion* ranks, and -0.15 for the ranks of users' ratings of their opinion of the colours and their *Extraversion*. For the Extravert interface, the resulting r_s values were 0.02 for the ranks of users' ratings of their liking of the interface and their percentage extraversion and -0.17 for the ranks of users' ratings of their liking of the colours and their *extraversion*. The small negative r_s values calculated for the Introvert interface were surprising, as we had expected positive r_s values, regardless of the magnitude of the values.

We determined the significance of these correlations using the t-distribution tables. This was because the Spearman table of critical values only catered for sample sizes up to 30. Since we considered the responses of the individual participants, our samples exceeded this value. For 30 - 40 degrees of freedom (dof = sample size minus 1), the critical t-statistic value was 2.042 for $\alpha = 0.05$ (two-tailed test). Since the t-statistic values computed for Introvert Memory liking and colours were -1.01 and -0.87, and the t-statistic values computed for Extravert Memory Liking and Colours were 0.14 and -1.01, we can say that none of these correlations were significant.

We then plotted the participants' opinion-scale responses on a histogram. As previously stated in Chapter 4, we decided to use 50% as the cut-off value when grouping users as *introvert* or *extravert* (i.e. users that have a level of *Extraversion* between or equal to 0% and 49% were deemed *introvert*, while users that scored 50% to 100% inclusive were deemed *extravert*). While it is arguable that people with a level of *Extraversion* around 50% could show characteristics of either *extravert* or *introvert* people, we felt that people with a level of *Extraversion* greater than 50% would display more defined *Extraversion* characteristics than those below 50%. From Figure 5.7, we can see that most users actually liked both system interfaces, since a rating of 1-3 corresponded to "Strongly Agree", "Quite Agree" and "Little Agree". We can also see that *extravert* participants had more extreme opinions - more *extravert* participants gave a rating of 1 or 7 to the interface.

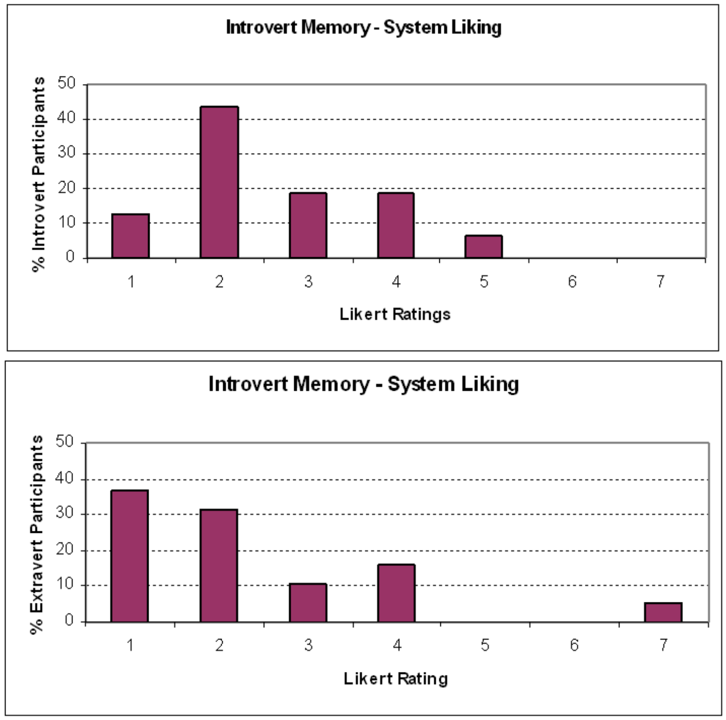


Figure 5.7: *Opinion-scale ratings from both* extravert *and* introvert *participants on the Introvert Memory interface*

For instance, 37% of *extravert* participants gave the Introvert interface a rating of 1, in comparison to 13% of *introvert* participants who gave the same rating, while 5% of *extravert* participants gave the interface a rating of 7, while none of the *introvert* participants gave this rating. We can also see a similar trend for ratings on the Extravert Interface (see Figure 5.8). 25% of *introvert* users and 23% of *extravert* users gave a neutral or dislike rating (i.e . 4 - 7) to both interfaces. In terms of comments passed, some of the Extraverts who like the *Extraverted* interface made comments that reflected their personality in terms of system initiated behaviour e.g. Dyad 2 User 2, who was highly extraverted, stated “It will narrow the options automatically rather than pressing the Hints button every time”. Similarly, an introverted user, Dyad 9 User 2 stated “If you are trailing behind in the game, the hints button is a useful handicap on the other player. However, it is a fairer test of memory it”. The rest of the comments made with regards to these interfaces can be found in Appendix A.

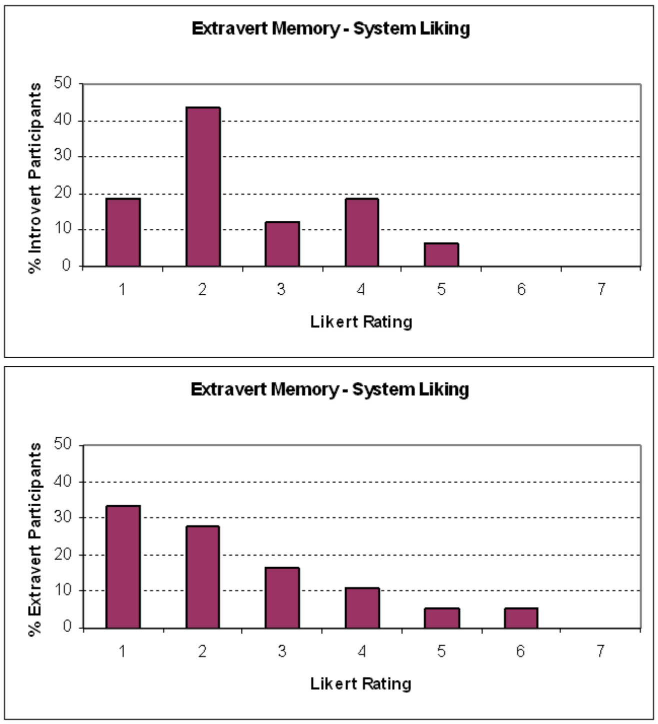


Figure 5.8: *Opinion-scale ratings from both* extravert *and* introvert *participants on the Extravert Memory interface*

From Figure 5.9, we see that both *extravert* and *introvert* participants gave quite similar ratings to the colours on the Introvert interface. 12.5% of *introvert* participants gave a neutral or dislike rating for the colours on this interface, while 10% of *extravert* participants gave a neutral or dislike rating. We had actually anticipated that more *extravert* participants would dislike these interface colours than *introvert* participants.

Looking at Figure 5.10, we see that 38% of *introvert* participants gave a neutral or dislike rating to the colours of the Extravert interface, in contrast to just 20% *extraverts* - a result we had anticipated i.e. more *extravert* users would prefer the Extravert interface than *introvert* users.

Finally, we compared the ratings of each participant for both interface liking

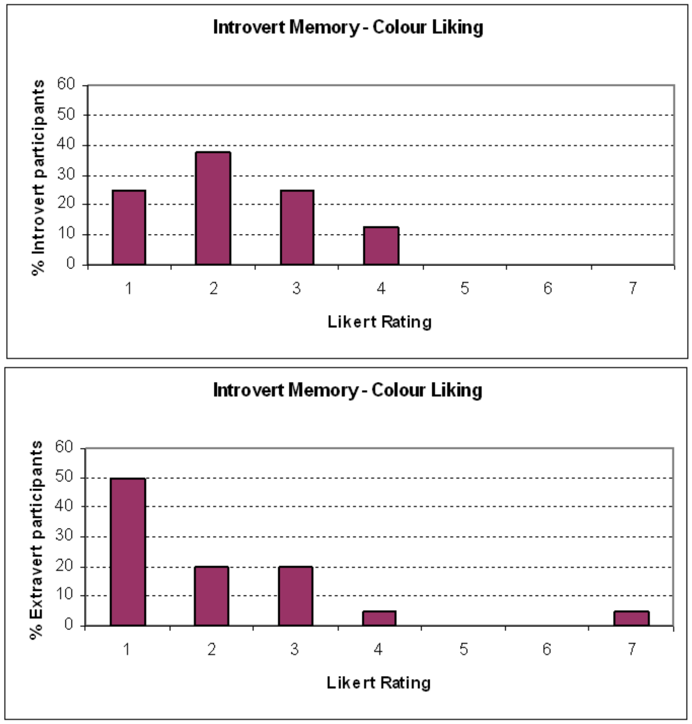


Figure 5.9: *Opinion-scale ratings on the colours of the Introvert Memory interface*

and colours, in order to see which interface each user preferred. The results of this are shown in Tables 5.11 and 5.12. We can see from Table 5.11 that in general, more people preferred the Introvert interface over the Extravert interface. However, it is interesting to see that more *introvert* users rated the Extravert interface higher than the Introvert interface, and more *extravert* users rated the Introvert interface higher than the Extravert interface - a result we certainly weren't expecting. In terms of interface colours, Table 5.12 shows us that all users had an overwhelming preference for the colours on the Introvert interface.

	Introvert Users	**Extravert Users**	**Total**
Introvert Interface	7	11	18
Extravert Interface	8	8	16

Table 5.11: *Interface preferences for Competitive Memory*

One should note that not all users' responses were included in these tables.

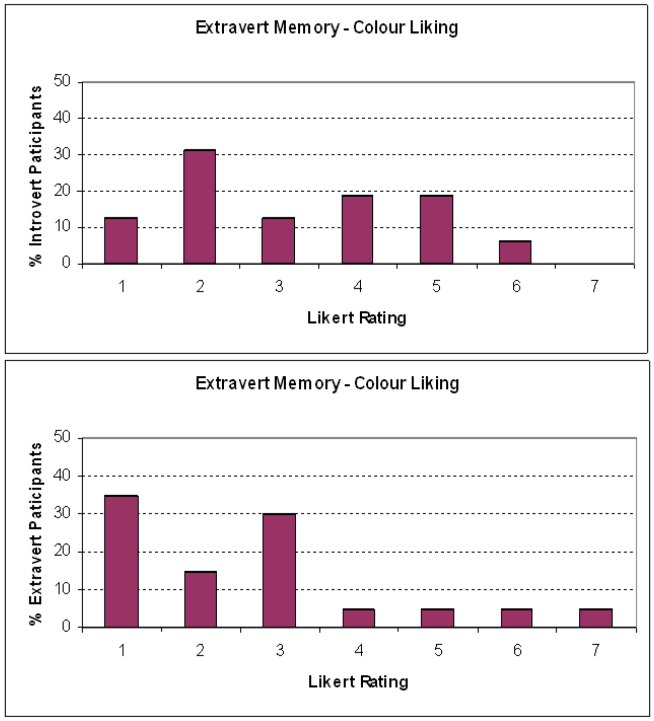


Figure 5.10: *Opinion-scale Ratings on the Colours of the Extravert Memory interface*

	Introvert Users	Extravert Users	Total
Introvert Interface	5	6	11
Extravert Interface	1	1	2

Table 5.12: *Colour preferences for Competitive Memory*

This was due to the failure of some users to provide a rating, or users giving the same rating to both interface variants. Since having the same rating indicated no preference, we did not include their responses in these tables.

Next, we looked at the interfaces of our Físchlár-DT 2 system. We looked at the opinion-scale ratings of each participant and compared their individual ratings to their respective scores on their *extraversion* personality trait. We again used a Spearman correlation in order to identify whether a correlation existed between these ranked variables (i.e. ratings and level of *Extraversion*) for each interface.

Once again, we anticipated a positive correlation between increasing *ex-*

traversion and increasing opinion-scale ratings on the Introvert interface and a negative correlation between increasing *Extraversion* and increasing opinion-scale ratings on the Extravert interface. We found that the correlation coefficients (r_s) calculated for this system were greater in magnitude than those calculated for the Competitive Memory system. We found a small positive correlation between users' opinion of the Extravert interface (r_s = 0.25) and negative correlations for layout and colours of this interface (r_s = -0.31 and -0.15 respectively). The negative correlation between increasing opinion-scale ratings for layout and increasing *extraversion* is logical, since we expected that *introvert* users would express a disliking of the interface. This was because the layout of the Extravert interface forced people to engage with each other and hence made the system more social.

We found negative correlations for the users' opinion of the Introvert interface, as well as the layout (r_s = -0.34 and -0.12 respectively) and a very small positive correlation for the colours used on the interface (r_s = 0.09). The negative correlation for interface liking and ranks of *Extraversion* is surprising, as it indicates that more extravert users preferred the Introvert interface.

Again, we looked at the t-statistics calculated to determine the significance of these correlations. Our critical value remained the same here as that used in the Competitive Memory system (2.042). The t-statistics calculated on the Introvert interface for liking, layout and colour were -1.96, -0.69 and 0.52 respectively. On the Extravert interface, the t-statistic values were 1.42, -1.88 and -0.88 for liking, layout and colour, respectively. Hence, this shows that the correlations calculated were not significant.

Next, we graphed the opinion-scale ratings for each interface in histograms. From Figure 5.11 we can see that none of the *extravert* users expressed a dislike of the Introvert interface, while 6% of *introvert* users expressed a dislike of the interface. In contrast, Figure 5.12 shows that no *introvert* users disliked the Extravert interface, while 30% of *extravert* users disliked the Extravert interface. This was very surprising as it meant that users preferred to work on the system

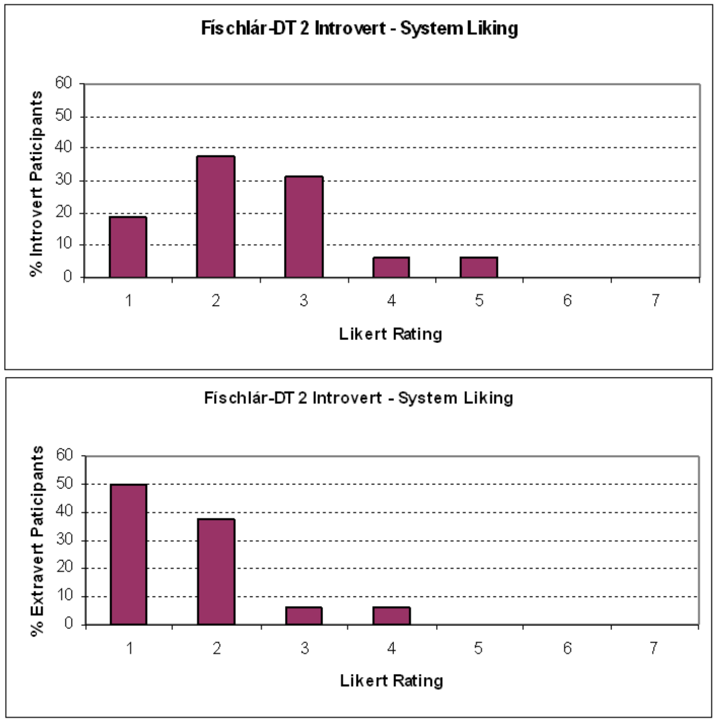


Figure 5.11: *Opinion-scale ratings on the Introvert Físchlár-DT 2 interface*

interface that was opposite to them in terms of their level of *Extraversion*.

Looking to the comments made by users, we see that some users did pass comments that reflected their personality types. For instance, Dyad 9 User 2, who was a more introverted person stated "I liked the separate functions better because it allowed us to work more efficiently, whilst still concentrating on the common goal". Similarly Dyad 2 User 1, who was also introverted stated "Separate Functions makes my action easily done". Dyad 14 User 2, who was an extraverted person stated that he liked "The layout and colours in the Shared Functions. The independence allowed by separate controls". Many other comments passed did not reflect a particular personality type, but more made general comments like Dyad 2 User 2's comment about the Extraverted interface: "It's user friendly, compact, gives you bigger area to view the pictures". The rest of these post-experiment questionnaire responses can be shown in Appendix A.

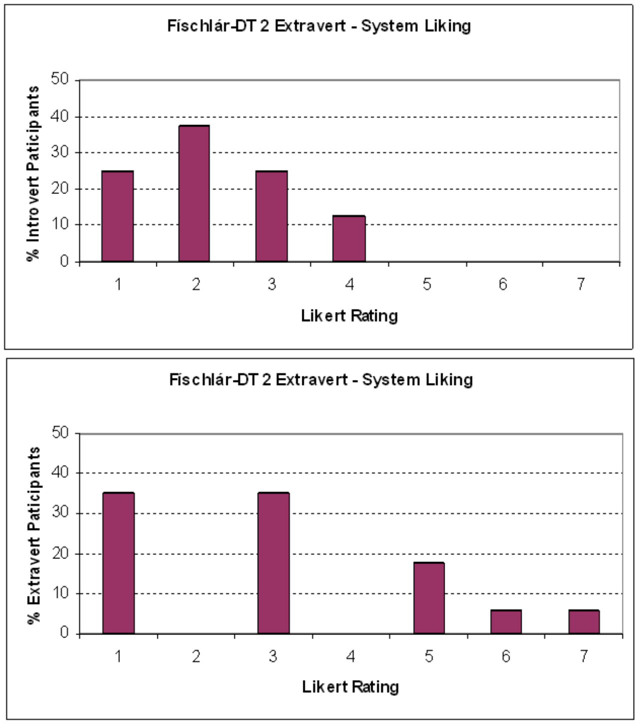


Figure 5.12: *Opinion-scale ratings on the Extravert Físchlár-DT 2 interface*

Looking at Figure 5.13, we can see that 25% of both *introvert* and *extravert* users had a neutral or dislike rating of the Introvert interface colours, with more *introvert* users disliking the colours (12%) than *extravert* users (10%). We had expected that the opposite would be the case. From Figure 5.14, we see that 50% of *introvert* users had a neutral or dislike rating of the Extravert interface colours, while 20% of the *extravert* users had a neutral or dislike rating of the interface colours - a finding that we would have anticipated.

Our graphs in Figure 5.15, show that *extravert* users had a more extreme disliking of the layout of the Introvert interface than the *introvert* users, with 5% of users giving it a "Strongly Disagree" rating. 25% of *introvert* users gave the layout a neutral or dislike rating, while 20% of *extravert* users gave it a neutral or dislike rating.

Figure 5.16 shows that no *extravert* users disliked the layout of the Extravert interface, while 13% of *introvert* users disliked the layout. 19% of *introvert* users

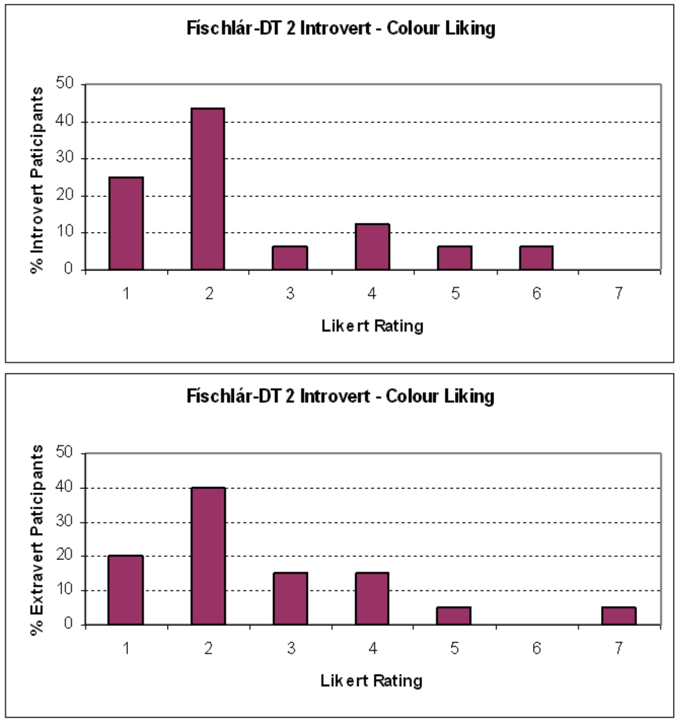


Figure 5.13: *Opinion-scale ratings on the colours of the Introvert Físchlár-DT 2 interface*

gave a neutral rating, in comparison to just 5% of *extravert* users who gave the same rating. These results were in line with our expectations.

Looking now to preferences, we see that once again, more users, in general, preferred the Introvert interface to the Extravert interface. Greater numbers of both *introvert* and *extravert* users expressed a preference for the Introvert interface. In terms of layout, more *introvert* users preferred the Introvert interface, while more *extravert* users preferred the Extravert interface. Finally, looking at users' colour preferences, we see that an equal number of *extravert* users preferred the Introvert and Extravert interface, while one more *introvert* user preferred the Introvert interface.

These findings are all summarised in Tables 5.13, 5.15 and 5.14

From the results above, we can see that the preferences of users are very mixed. It appears that users liked some of the features that reflected their level

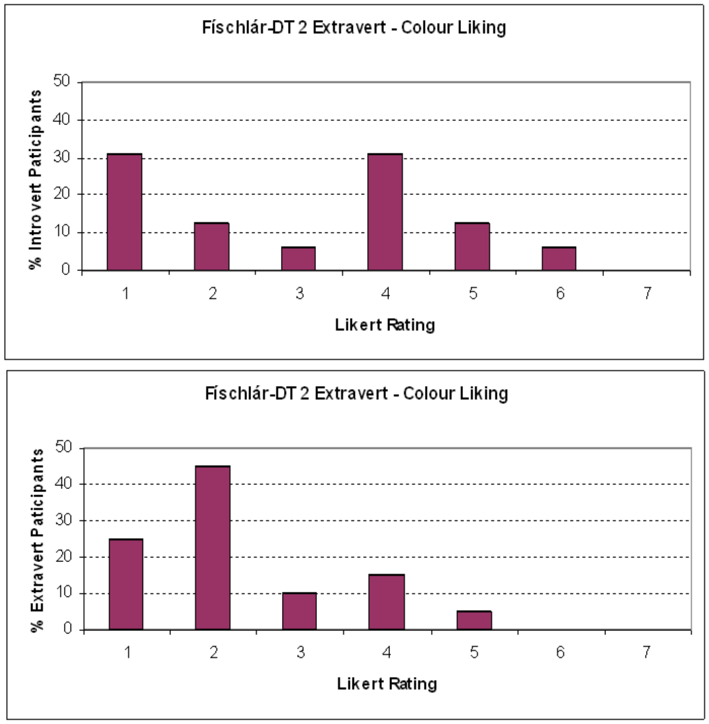


Figure 5.14: *Opinion-scale ratings on the colours of the Introvert Físchlár-DT 2 interface*

	Introvert Users	**Extravert Users**	**Total**
Introvert Interface	8	14	22
Extravert Interface	7	6	13

Table 5.13: *Interface preferences for Físchlár-DT 2*

of *extraversion*. This is true of the layout of the Físchlár-DT 2 system, and somewhat true of the colours used in the interface variants of this system. In the case of the interface variant preferences for the Competitive Memory system, the opposite was actually true. For the remaining aspects (colour preferences on the Competitive Memory system and interface variant preference on the Físchlár-DT 2 system), both sets of users showed a majority preference for the

	Introvert Users	**Extravert Users**	**Total**
Introvert Interface	6	4	10
Extravert Interface	3	6	9

Table 5.14: *Layout preferences for Físchlár-DT 2*

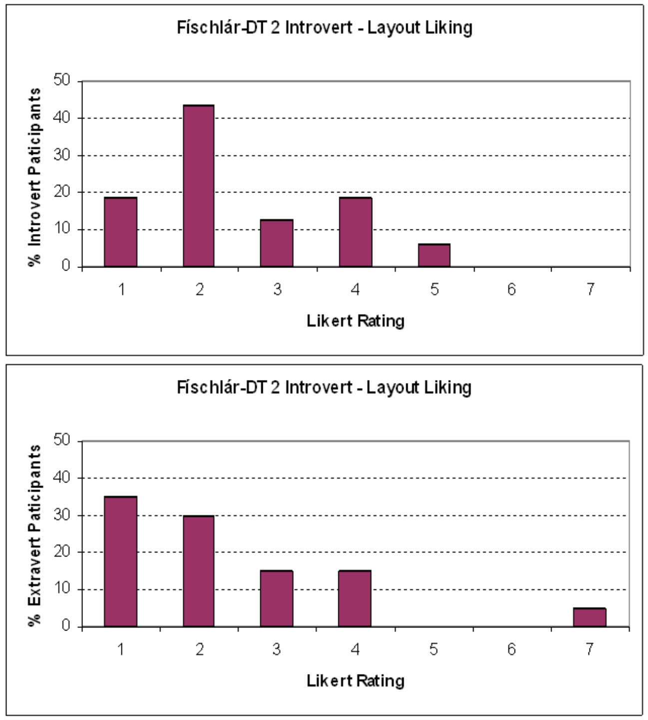


Figure 5.15: *Opinion-scale ratings on the layout of the Introvert Físchlár-DT 2 interface*

Introvert system.

	Introvert Users	**Extravert Users**	**Total**
Introvert Interface	5	5	10
Extravert Interface	4	5	9

Table 5.15: *Colour preferences for Físchlár-DT 2*

We also noted that very few people actually expressed a disliking of any of the interfaces. Most users either liked all system variations, or occasionally gave a neutral response. Most users' preferences were only differences of one or two opinion-scale points. Hence, users did not express extreme preferences of one interface over another, simply mild ones.

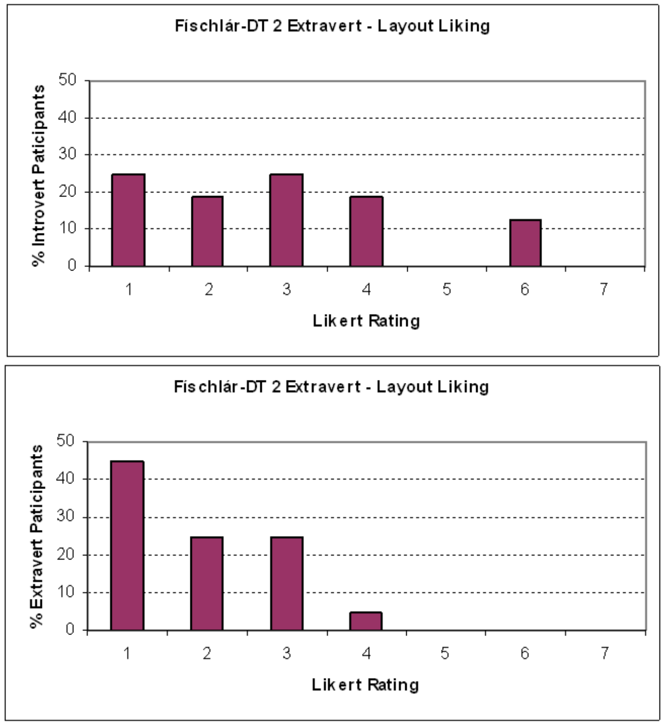


Figure 5.16: *Opinion-scale ratings on the layout of the Extravert Físchlár-DT 2 interface*

In conclusion, we can see that users do not always prefer aspects of an interface that match their level of *extraversion*. This includes the overall system ratings, colours in our Competitive Memory interface variants, as well as overall system rating in our Físchlár-DT 2 interface variants. However, this is more apparent in the case of the *extravert* users.

Q 7. Do dyads perform better on an interface variant/under a task constraint variant that they like better, when give two variants ?

Our prediction here was that there would be a relationship between a dyad's performance and their interface/constraint preference. We expected that if the users enjoyed working on an interface, they would become more enthusiastic about doing the task and so this would have a positive effect on their performance. To quantitatively identify if this was in fact the case in our experiments, we looked at the users' post-experiment questionnaire data and their performance results. We examined only our collaborative systems, since our competitive systems require that users work on their own in order to defeat their opponent. Hence, there is no "dyad preference" in competitive systems.

We look first to our **Collaborative Memory** System. Table 5.16 contains the participating dyads, listed in order of decreasing performance for each of the rules of the system. We recall that the two rules imposed on this system were (1) to complete the game with as few mismatched pairs as possible (Accuracy Memory) and (2) to complete the game as quickly as possible (Speed Memory). A 'Y' is placed in the column labeled "Both Like" for a system rule if both members of a dyad mutually liked the rule. If dyads liked both rules equally, then a 'B' is placed in the "Both Like" column for both rules.

Looking at the 'Y's, we notice that more dyads preferred the Accuracy version of the system than the Speed version. Of the six dyads that expressed a like for the Accuracy version, four performed better under that rule than in the Speed counterpart. Two dyads performed worse. In fact, Dyad 10 ranked first place with their performance on the Speed version of the system, even though they preferred the Accuracy version. The two dyads that preferred the Speed version performed better on this version than on the Accuracy version in comparison to the other dyads. Hence, six out of eight dyads (i.e. 75%) performed better on the interface that they preferred.

The **Físchlár-DT 1** system also gave us a result that supported our predic-

Accuracy Rule			Speed Rule		
Dyad #	**Mismatches**	**Both Like?**	**Dyad #**	**Time Taken (in seconds)**	**Both Like?**
Dyad 17	38		**Dyad 10**	212	
Dyad 1	43	B	**Dyad 7**	291	Y
Dyad 15	45	Y	**Dyad 1**	302	B
Dyad 8	47	Y	**Dyad 12**	312	
Dyad 9	47	Y	**Dyad 14**	321	
Dyad 6	53	Y	**Dyad 15**	321	
Dyad 12	55		**Dyad 16**	325	Y
Dyad 10	56	Y	**Dyad 8**	334	
Dyad 7	59		**Dyad 18**	335	
Dyad 14	59		**Dyad 6**	336	
Dyad 18	60		**Dyad 17**	341	
Dyad 4	66		**Dyad 3**	357	
Dyad 11	68	B	**Dyad 11**	363	B
Dyad 16	68		**Dyad 5**	369	
Dyad 2	72		**Dyad 13**	378	B
Dyad 3	85	Y	**Dyad 4**	386	
Dyad 5	99		**Dyad 2**	387	
Dyad 13	102	B	**Dyad 9**	405	

Table 5.16: *Dyads' performance and preferences on Collaborative Memory*

tion. Here, we looked at whether people liked the system or not and how this related to their performance under both rules. From Question 5 above, we saw that nine dyads expressed a mutual agreement in their opinion of the system, eight of whom mutually liked the system, with the remaining dyad disliking the system (Dyad 8).

Table 5.17 below shows dyads listed in order of decreasing performance for each of the rules of the system. Dyads that liked the system contain a 'Y' in the column labeled "Both Like". As we can see, dyads that mutually liked the system generally performed better for both rules, but in particular the *10 Minute Rule.* When looking at this table, we must take into consideration that Dyads 2 and 6 worked in an information retrieval environment at the time that these experiments were carried out (which we learned from our pre-experiment questionnaire data), and so would be accustomed to working with various types of search systems.

10 Minute Rule			Find 10 Rule		
Dyad #	**Recall (# shots)**	**Both Like?**	**Dyad #**	**Time Taken (in seconds)**	**Both Like?**
Dyad 18	54	Y	**Dyad 17**	312	Y
Dyad 6	47		**Dyad 1**	355	Y
Dyad 2	39		**Dyad 6**	405	
Dyad 3	32	Y	**Dyad 18**	417	Y
Dyad 4	29	Y	**Dyad 4**	420	Y
Dyad 10	27	Y	**Dyad 16**	438	Y
Dyad 16	27	Y	**Dyad 11**	554	
Dyad 5	26	Y	**Dyad 7**	570	
Dyad 1	25	Y	**Dyad 2**	595	
Dyad 11	25		**Dyad 10**	610	Y
Dyad 7	24		**Dyad 3**	611	Y
Dyad 13	17		**Dyad 5**	625	Y
Dyad 15	16		**Dyad 9**	690	
Dyad 8	15		**Dyad 13**	690	
Dyad 9	15		**Dyad 8**	795	
Dyad 17	14	Y	**Dyad 15**	822	
Dyad 12	6		**Dyad 14**	1188	
Dyad 14	3		**Dyad 12**	1207	

Table 5.17: *Dyads' performance and preferences on Físchlár-DT 1*

In our **Físchlár-DT 2** system, we noted that fewer dyads preferred the Extravert interface than the Introvert interface. Table 5.18 displays dyads in order of decreasing performance on the Extravert and Introvert interfaces. Two dyads stated a preference for the Extravert interface and we can see from the table that they did in fact perform better on that interface than on the Introvert interface in comparison to the other dyads. Eight dyads expressed a preference for the Introvert interface – however, three of these actually performed worse on the Introvert interface, moving down one, two and seven performance rank positions respectively. Since seven out of ten dyads' performances actually ranked (70%) more highly in terms of performance when they worked on the interface they liked better, we can deduce that generally the dyads worked better on the interface that they preferred.

In conclusion, the majority of dyads performed better on the system interface or constraint that they preferred.

Extravert Interface			Introvert Interface		
Dyad #	**Recall (# shots)**	**Both Like?**	**Dyad #**	**Recall (# shots)**	**Both Like?**
Dyad 5	41	Y	**Dyad 18**	33	Y
Dyad 7	41		**Dyad 13**	27	
Dyad 6	40	Y	**Dyad 17**	27	
Dyad 17	39		**Dyad 5**	22	
Dyad 2	37		**Dyad 3**	19	Y
Dyad 18	37		**Dyad 6**	18	
Dyad 4	29		**Dyad 15**	18	Y
Dyad 16	29		**Dyad 2**	16	
Dyad 10	28		**Dyad 7**	14	
Dyad 8	27		**Dyad 10**	13	Y
Dyad 11	23		**Dyad 12**	13	
Dyad 12	22		**Dyad 1**	11	Y
Dyad 3	21		**Dyad 11**	11	Y
Dyad 14	20		**Dyad 4**	10	Y
Dyad 1	19		**Dyad 8**	10	
Dyad 13	11		**Dyad 14**	8	
Dyad 15	10		**Dyad 9**	6	Y
Dyad 9	8		**Dyad 16**	5	

Table 5.18: *Dyads' performance and preferences on Físchlár-DT 2*

Q 8. Is there a relationship between a user's stated opinions on a system and their interaction data ?

For this question, we looked at the responses with regards to logged interaction data (i.e. our CCTV footage) and touch-point data that we gathered from each of our participants individually. This was in order to see if the participants' opinions of each interface, as denoted by their ratings along the 7-point opinion-scale, were correlated to their interactions and touch-points.

Firstly, we looked at the interaction instances (as described along our taxonomy of interaction instances in Section 5.1.3 above i.e. the total of each dyad's requests, responses and comments) of each individual and observed if those with more interaction instances liked or disliked each of the systems. We noted the ratings each participant gave for the system interfaces and task constraints and their respective interaction instances. To identify whether these were correlated, we carried out a Spearman rank correlation between the two variables. The resulting correlation coefficients are given in Table 5.19. We did not examine dyads' interaction instances for the *Pop-A-Bubble* game, since so few dyads actually communicated during any of the games.

We had different sample sizes for each of the systems (due to some corrupted videos) for all but the Introvert and Extravert Memory game interfaces and also incomplete questionnaire responses, where people did not rate their liking of one of the interfaces. As a result, our n values for Accuracy and Speed Memory were both 34; for Introvert and Extravert Memory, they were 35 and 34 respectively; for Físchlár-DT 1 (10 min) and (Find 10) rules, they were both 34 and for Físchlár-DT 2 Extravert and Introvert interfaces, the n values were 29 and 27 respectively. Since our n values were different, we had different critical values for each system. Also, since the Spearman correlation tables only cater for small sample sizes (i.e. less than 30), we were required to look at the t-statistics computed for these samples. Hence, for all but the Físchlár-DT 2 variations, we looks at the t-statistics.

	r_s / t-stat.	Critical Values	Significant?
Accuracy Memory	-0.87	2.042	No
Speed Memory	-0.25	2.042	No
Introvert Memory	-0.62	2.042	No
Extravert Memory	1.72	2.042	No
Físchlár-DT 1 (10 min)	0.71	2.042	No
Físchlár-DT 1 (Find 10)	0.63	2.042	No
Físchlár-DT 2 (Int)	-0.48	0.392	Yes
Físchlár-DT 2 (Ext)	0.27	0.377	No

Table 5.19: *Spearman rank correlation of system opinion-scale ratings and interaction instances*

The associated critical values for each of the systems are supplied in Table 5.19, along with a column indicating whether the correlation coefficients computed were statistically significant or not.

From this table, we can see that just the Físchlár-DT 2 Introvert interface shows a strong negative correlation between interaction instances and the participant's liking of the interface. Here, participants who liked the interface more, i.e. had low opinion-scale ratings, communicated more on the interface. Since participants were forced to interact more on this system, i.e. they had separate resources and so had to coordinate their actions in order to work together sucessfully, it is unsurprising that they communicated more.

Since users had their own set of buttons on the Físchlár-DT 2 Introvert interface, they had to coordinate their actions better in order to avoid playing over each other's videos or covering over each other's search results. Those who didn't communicate well would have experienced more frustration at having their actions over-ridden by their partners and hence, would dislike this interface as a result. The other Físchlár-DT interfaces had very similar function layouts, hence they had similar correlation values, i.e. small positive correlation coefficients.

We also looked at coordination errors for the Físchlár-DT 1 and 2 systems. We carried out a Spearman rank correlation for opinion-scale ratings and coordination errors. Our resulting correlation coefficients were 0.13 and -0.04 for

	t-statistic	Critical Values	Significant?
Accuracy Memory	0.6	2.042	No
Speed Memory	-0.42	2.042	No
Introvert Memory	-0.25	2.042	No
Extravert Memory	0.73	2.042	No
Físchlár-DT 1	-1.78	2.042	No
Físchlár-DT 2 (Int)	0.74	2.042	No
Físchlár-DT 2 (Ext)	-1.11	2.042	No

Table 5.20: *Resulting t-statistics of system opinion-scale ratings and individual touch-points*

Físchlár-DT 2 (Extravert interface) and Físchlár-DT 2 (Introvert interface) respectively. For Físchlár-DT 1 (10 minute rule) and Físchlár-DT 1 (Find 10 rule), we had to compare the t-statistic values to the critical value (2.042 at α = 0.05, two-tailed test). These values -0.22 and -0.20 respectively - statistically insignificant results. The critical values differed for Físchlár-DT 2 Introvert and Extravert interfaces - 0.392 and 0.377 respectively. Hence, none of these correlation coefficients indicated a significant relationship between coordination errors and opinion-scale ratings of the systems.

We applied the same approach in examining the touch-points and opinion-scale ratings given from all participants. Table 5.20 displays the correlation coefficients obtained for opinion-scale ratings and touch-points for each of our systems. Once again, the n values differed for each of the systems, due to a lack of ratings supplied by some of the participants. Hence, our n values for Accuracy and Speed Memory were both 36; for Introvert and Extravert Memory, the n values were 35 and 34; for Físchlár-DT 1, our n value was 36 and finally for, Físchlár-DT 2 Extravert and Introvert interfaces, our n values were 33 and 32 respectively. Hence, we used the t-statistic values for all our systems in this instance.

From Table 5.20, we can see that there are no significant correlations between participants' touch-points and their opinion-scale ratings of the systems.

The post-experiment questionnaire for Pop-A-Bubble did not elicit opinion-scale ratings from participants regarding their opinion of the system, but rather they indicated the version of the system that they preferred. To see if their preference was related to their touch-points, we ranked the interfaces by increasing number of touch-points for each interface (e.g. if a user had 200, 250 and 190 touch-points on the Dual Track-bar, Single Track-bar (You Control) and Single Track-bar (Other controls), then these ratings would be 2, 3 and 1 respectively). We then noted the touch-point rating of the interface that they preferred in order to see if users generally had more or less touch-points on the interface they preferred. Results from 28 participants were examined (the remaining 8 either noticed no difference and hence did not select an interface preference, or they were members of the 3 dyads that contained a colour-blind member).

The results showed that 43% had the lowest number of touch-points on their preferred interface (i.e. a rating of 1), 39% had a touch-point rating of 2 on their preferred interface and 18% had the highest number of touch-points on their preferred interface. This outcome suggests that users touched the table less often on the interface variation they preferred.

From the above, we can conclude that logged and systematically recorded interaction data does not generally relate to users' stated opinions on a system, with the exception of the Físchlár-DT 2 Introvert interface, which statistically showed that those who liked the system, communicated more. Our Pop-A-Bubble system also displayed a tendency for users to physically interact with the tabletop's surface less frequently on the interface that they preferred.

Hypothesis Support / Disproval

From looking at the answers to our sub-questions, we can see that dyads did not intuitively prefer the interfaces that reflected their level of *Extraversion* as we had expected. Members of dyads had similar opinions on their interface preferences just over half of the time and their preferences did not always relate to their personality type. However, we did find that dyads generally did work better on an interface that the members collectively preferred. Hence, our findings only support the second aspect of this hypothesis i.e. we cannot say that dyads will prefer interfaces that reflect their personality types in appearance and behaviour (in our case, personality match between interface and dyads is primarily based on the *Extraversion* personality trait); but we can say that dyads work better on interfaces/under task constraints that they prefer.

5.5 Hypothesis 3

Dyads perform different tasks in a different manner and this is related to their personality.

This hypothesis is much more general and as we can see from Section 5.2, it raises a greater number of interesting hypothesis sub-questions. We now proceed to answer these.

Q 9. How does imposing different constraints on a collaborative task affect the performances of the dyads ?

To answer this, we look at the performances of the dyads on our Collaborative Memory Game and Físchlár-DT 1. To recap, Collaborative Memory Game imposed two different rules for the same four games. The first rule required dyads to complete all four games as accurately as possible i.e. with the least number of mismatched pairs. The second rule insisted that dyads complete all four games as quickly as possible, regardless of how many mismatched pairs they uncovered. However, the end goal was the same for both rules - to collaboratively find all the matching pairs of cards. While these were different constraints, the overall goal was the same - to match all pairs of cards collaboratively.

For these two rules, we recorded both the time taken to complete the games and the number of mismatched pairs, to see if significant differences arose between the *Accuracy* rule and the *Speed* rule. The overall average number of mismatches for all games played under the *Accuracy* rule was 14.64, with a standard deviation of 4. The same figure for all games played under the *Speed* rule was 18.13 (with a standard deviation of 3), an average increase of almost 3.5 mismatches. Similarly, the average time taken to complete the games played under the *Speed* rule was 1 minute and 24 seconds, with a standard deviation of 11 seconds. The same figure for the *Accuracy* rule was 1 minute and 47 seconds (with a standard deviation of 17 seconds) - an increase of 23 seconds (or 27%).

To test whether these increases were statistically significant, we carried out a Student T-Test (Paired Two Sample for Means) for the number of mismatches under each rule and the time taken to complete the games under each rule. For n-1 (17) degrees of freedom, and an α of 0.01, our t-statistic for the correlation of the number of mismatches under both conditions was -5.09, with a critical value of 2.90, indicating that the two means are highly significantly different. We obtained a similar outcome for time taken under both rules - our t-statistic was 4.20, indicating a significant difference in the mean times taken under each rule.

We also look at the performance of the dyads to see if the same dyads score highly on both systems. Looking back at Table 5.16 from Question 7 above, we see the ordering of decreasing dyad performance for each rule. One can see that the dyad orderings are rather different. Only 1 dyad (Dyad 11) has the same performance rank under both Accuracy and Speed rules, while Dyad 1 had just a difference of 1 in both performance rankings. This indicates that some dyads perform better under certain constraints than others.

For Físchlár-DT 1, we look only at the performance rankings. We do not look at the performance of dyads in terms of productivity because the topics searched for are not consistent in their level of difficulty across the two rules. Hence, we cannot compare actual performances. Instead, we look to the rankings of the dyads performed across both rules (i.e. relative performance).

Looking back at Table 5.17, we see that there is a slightly greater similarity between the dyads' performance ranks for both systems. Dyad 4 had the same performance rank for both rules, while Dyads 16, 8, 12 and 14 had just a difference of 1 rank position. However, all other dyads displayed a greater variation in their performance ranks.

Conclusion: Imposing different rules or constraints on a system task does significantly impact on the performance of a dyad. In the case of time and accuracy, the data gathered from our dyads has implied that imposing one of these as a constraint on a collaborative card game, results in a significant deterioration in the other. For a collaborative search task, relative performance is generally different for dyads when working under the different constraints that we imposed.

Q 10. Are there more interaction instances in a collaborative version of a game as opposed to a competitive version ?

While this is not directly related to our third hypothesis, we believed it would be interesting to look at the communication of dyads across collaborative and competitive versions of a task, to determine if dyads who communicate a lot in a collaborative task, still communicate a lot in a competitive task. Since communication potentially affects how a task is performed, we chose to place this sub-question here. In order to answer this question, we compare the collaborative and competitive versions of Memory Game. For all dyads (with the exception of Dyad 5 as there was a fault with the video recording for their competitive games) and all games played on the competitive version of Memory Game, the total number of interaction instances was 827. The same figure for the collaborative version of Memory Game was a 2,910 - an increase of 252%. We had expected that there would certainly be more interaction instances in the collaborative version, as naturally people communicate in a task that they have to work together on. When playing a competitive game, one tries to outsmart their opponent, and so communication is kept to a minimum in case the opponent learns of one's intentions or tactics.

Figure 5.17 shows the total amount of interaction instances for each dyad on each system. One can see that all dyads had more interaction instances on the collaborative version of the system than the competitive version. Figure 5.18 illustrates the relative percentage increase of interaction instances between the Competitive Memory Game and Collaborative Memory Game. What is interesting to note here is that the three dyads with the least relative increase in interaction instances were actually *Extravert/Extravert* pairs (i.e. where both members of the dyads had a score greater than or equal to 50% on the *Extraversion* personality trait).

Dyad 15, Dyad 7, Dyad 2, Dyad 16 and Dyad 12 had the greatest relative increase in interaction instances. Of these five dyads, two were *Introvert/Introvert*

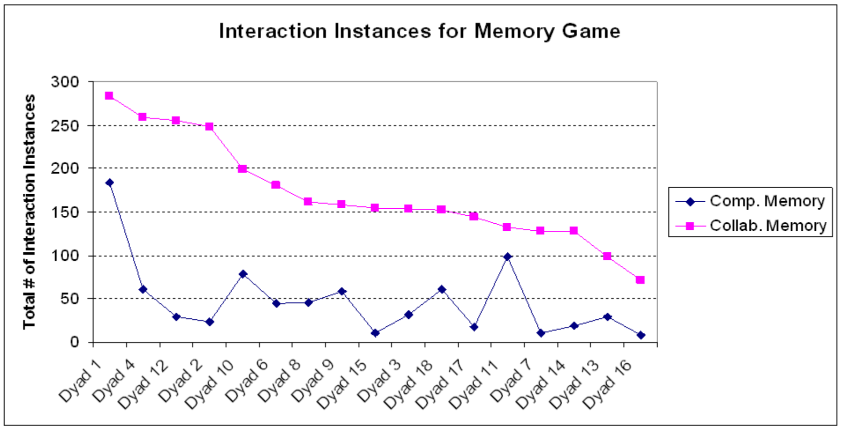


Figure 5.17: *Comparison of interaction instances for Competitive and Collaborative Memory*

dyads, two were *Extravert/Introvert* dyads and one was an *Extravert/Extravert* dyad. Dyads 9, 10, 18, 1 and 11 had the smallest relative increase in interaction instances. Of these five dyads, three were *Extravert/Extravert* dyads and two were *Extravert/Introvert.*

These results lead us to believe that dyads with *introvert* members displayed a greater relative difference in their instances of interaction between the competitive and collaborative versions of the system.

Conclusion: Dyads have more interaction instances in collaborative systems than in competitive systems.

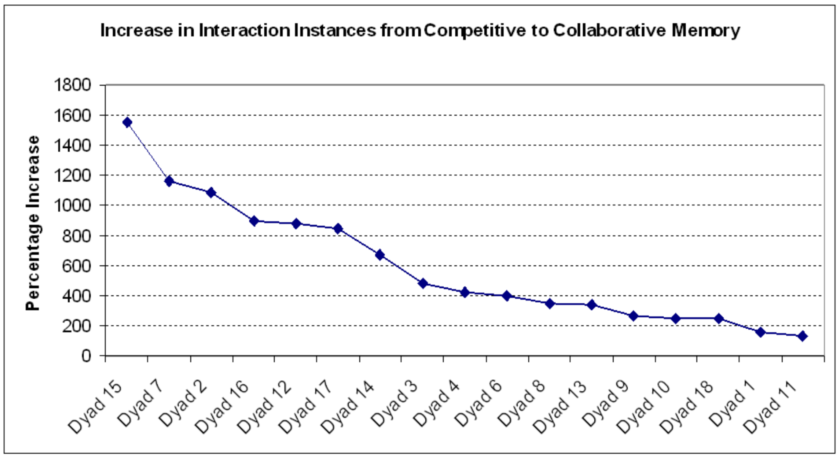


Figure 5.18: *Percentage increase in interaction instances for each dyad*

Q 11. Does the amount of interaction among a dyad relate to the performance of that dyad in our collaborative tasks ?

This question concerns the interaction instances annotated from the CCTV recordings of the experiments, which we attempted to correlate with the performance of the dyads. Not only this, but we also attempted to correlate the touch-point data of all dyads and their performance. We used the Spearman rank correlation method to correlate these 2 sets of data to dyads' performances. We chose only to do this for collaborative systems, since competitive systems did not have a "dyad performance".

Firstly, we looked at interaction instances and the participating dyads' performances to see if they were related. Using the Spearman rank correlation method involved creating a rank of performance positions in ascending order. We also created a rank of increasing interaction instances e.g. if Dyad 17 came 1st in terms of performance and 5th in terms of their number of interaction instances, we would place these side by side in our calculations.

	Spearman Correlation Coefficient (r_s)
Accuracy Memory	-0.36
Speed Memory	0.27
Físchlár-DT 1 (10 min)	0.08
Físchlár-DT 1 (Find 10)	0.45
Físchlár-DT 2 (Int)	-0.08
Físchlár-DT 2 (Ext)	-0.19

Table 5.21: *Spearman rank correlation of performance and interaction instances*

Table 5.21 contains the Spearman correlation coefficients (r_s values) calculated for the ranks. We can see from this table, that the r_s values obtained were very small for all but the Físchlár-DT 1 (Find 10 rule) system. Accuracy Memory had a small relationship between increasing performance and increasing interaction instances, while Speed Memory had a small relationship between decreasing performance and increasing interaction instances. Físchlár-DT 2 Introvert interface and Físchlár-DT 1 (10 minute rule) had almost zero r_s values, indicating that the relationship between performance and the number of interaction instances was almost completely random i.e. no correlation. Our Físchlár-DT 1 (Find 10) showed a positive relationship while a small negative r_s value was obtained for the Físchlár-DT 2 Extravert system.

From our statistical tables, we find that the critical values at $\alpha = 0.05$ (two-tailed test) are 0.507 for Accuracy Memory, Speed Memory, both Físchlár-DT 1 rules and for Físchlár-DT 2 Extravert. Físchlár-DT 2 Extravert had a critical value of 0.545 (see Question 1 above for details on the differences in critical values for each system). Hence, we see that none of our systems showed a significant correlation between interaction instances and performance on our collaborative systems. The r_s obtained for the Físchlár-DT 1 (Find 10) system was quite close to the critical value for this system, suggesting that those dyads who performed best had more overall instances of interaction (i.e. verbal and gestural communication) than those who performed worst. This would seem to be logical, since this was a highly collaborative task, so the more that each dyad communicated, the faster they would collaboratively find the 10 relevant shots required.

We would have anticipated a strong relationship between interaction instances and the Accuracy Memory system in particular, since it required strong collaboration between each dyad member. However, it is possible that people just ignored each other's suggestions, or that the card selector trusted the other person (match-finder) enough to allow them to chose themselves whatever card they believed to be the matching card. We must also remember that there was a significant relationship between dyad *Extraversion* and normalised interaction instances (see Question 1 above) for the Accuracy Memory, Físchlár-DT 1 (both rules), Físchlár-DT 2 Introvert and Físchlár-DT 2 Extravert systems, so this may have had an impact here, where just the participants' personalities affected their verbal and gestural interaction, but their performance was unrelated.

Since this was the total number of interaction instances, and since the length of time taken to find 10 shots varied greatly from dyad to dyad, we decided to look at the rate of interaction instances per minute, to see if this correlated to performance. We used one minute as our unit measure figure, since none of the dyads completed the task in less than one minute. Using our Spearman rank correlation method, we ranked each dyad's number of interactions per minute and their performance, and got an r_s value of -0.40 - a statistically insignificant inverse correlation.

We also did this for the Accuracy and Speed Memory Systems to see if there was a significant correlation between interaction instances per minute and performance. However, the resulting r_s values (-0.36 and 0.02 for Accuracy and Speed Memory respectively) also conveyed insignificant correlations. We did not attempt to correlate interactions per minute for the remaining systems, since they specified a time-frame within which all dyads had to complete the task. Hence, the resulting r_s values would not be different to those obtained above.

We thought it would be interesting to see if the number of coordination errors each dyad had was related in any way to their performance. One could anticipate that a higher number of coordination errors (i.e. actions that one user carries out that interrupt their partner's work), would result in lower levels of performance. Here, we looked only at the Físchlár-DT systems, since coordination was software imposed on the Memory Game systems.

Carrying out a Spearman correlation between performance on the Físchlár-DT 1 system for both rules, resulted in r_s values of 0.32 for the "10 minute" rule, and -0.57 for the "Find 10" rule. Since our critical value here was 0.507, the "Find 10" rule showed a significant inverse relationship between performance and coordination errors (i.e. those who have a higher performance on this system, have a lower number of coordination errors). This would appear to be a logical finding, since interaction errors cost valuable time - our measurement of success in this system. On a per-minute basis however, the r_s value was 0.40 – a statistically insignificant result. Our analysis of the Físchlár-DT 2 interfaces resulted in r_s values of 0.23 for the Introvert interface and 0.38 for the Extravert interface, that is, insignificant correlations.

We performed the same procedure to correlate the ranks of dyads' average performance data for each system to the average number of touch-points recorded for each system. Once again, two ranks were created of decreasing performance and number of touch-points, in an attempt to see if higher performance was related to a greater number of touch-points recorded. We had to combine performance data from both rules of the Físchlár-DT 1 system in order to compare against the system touch-points, since it was not possible to separate the touch-point data for each rule.

Table 5.22 shows the resulting Spearman correlation coefficients, r_s for the ranks obtained from each of the systems. Once again the r_s values obtained are small, though they are slightly larger in magnitude than those noted for the correlations between the performance and interaction instances variables above. Accuracy Memory has a very high r_s value, indicating a strong positive correlation between performance and touch-points. Físchlár-DT 2 Introvert interface, Físchlár-DT 2 Extravert interface and Físchlár-DT 1 (both rules) have relatively small relationships between performance and touch-points, while Speed

	Spearman Correlation Coefficient (r_s)
Accuracy Memory	0.95
Speed Memory	0.30
Físchlár-DT 1 (both rules)	-0.23
Físchlár-DT 2 (Int)	0.54
Físchlár-DT 2 (Ext)	0.42

Table 5.22: *Spearman rank correlation of performance and touch-points*

Memory has a small positive relationship between decreasing performance and increasing number of touch-points.

Our critical value here was 0.476, which shows that the Físchlár-DT 2 (Introverted interface) and Accuracy Memory game displayed significant relationships between dyad performance and touch-points. The Físchlár-DT 2 (Extraverted interface) gives an r_s value that is close to the critical value, suggesting that dyads that had more touch-points performed better, though this could not be statistically verified.

The strong correlation between touch-points and performance on our Accuracy Memory system was unsurprising, since it would appear logical that users who touched the tabletop less, performed better i.e. they were more accurate and finished the game with fewer attempts or mismatched pairs. One reason why we did not find a significant correlation for the Speed Memory system, could be that people were very conscious of the fact that they were under time pressure to complete the task and so didn't worry about how many cards they overturned. Instead, they selected cards more quickly (and frequently with less accuracy).

We would have assumed that the greater the number of touch-points dyads incurred on the Físchlár-DT 2 interfaces, the better they would have performed. While this is the case for the Introverted interface, the r_s value obtained for the Extraverted interface was just outside the critical value. However, this does suggest that users who have more touch-points are more inclined to perform better.

We then looked at the correlations between touch-points per minute and performance on the Accuracy Memory, Speed Memory and Físchlár-DT 1 systems. These gave r_s values of 0.50, -0.67 and -0.30 respectively. The significant correlation for Accuracy Memory is unsurprising, though the correlation on a per minute basis is weaker than that obtained for overall number of touch-points. However, we must note that time is not important in this system - being accurate is the most important factor. Speed Memory presented us with a surprisingly high significant inverse correlation, that is, better performing dyads had a lower number of touch-points per minute than the worse performing dyads. We can see that that the number of touch-points per minute is not significantly related to the performance of the dyad in the Físchlár-DT 1.

From the above, we can see that our systems and their respective interfaces did not produce a statistically significant correlation between overall interaction instances and performance. We found no correlation between interaction instances per minute and performance. However, we did find a significant inverse correlation between coordination errors and performance on the Físchlár-DT 1 (Find 10) system. For the Accuracy Memory system and Físchlár-DT 2 (Introvert interface), we found a significant correlation between the overall number of touch-points and performance. Looking at touch-points per minute and performance, we found significant correlations for Accuracy Memory and Speed Memory.

To conclude, we find that the amount of interaction among a dyad only relates to the performance of the dyad for some tasks, in our case, normalised touch-points affected performance on Accuracy Memory, Speed Memory and Físchlár-DT 2 (Introvert interface).

Q 12. Do dyads coordinate their actions well on our collaborative search tasks is this related to their personality type ?

This question relates to the number of coordination errors that each of the dyads incurred, principally for the Físchlár-DT systems (see Smeaton et al., 2007). To recap, coordination errors are errors that we annotated from the CCTV footage of our experiments, when one user's actions interrupted their partner's actions. The reason that we do not look at the coordination policy for the Competitive and Collaborative Memory Games is that coordination was imposed by the underlying software. Players had to take their turns as the software allowed e.g. if one player tried to overturn a card and it was not their turn, the system would not respond - it would only respond to the user whose turn it was.

We look briefly at the questionnaire results from our Pop-A-Bubble game - in the post-experiment questionnaires 16 out of the 18 participating dyads reported that their hands collided. One other dyad stated that they had no collisions and the remaining dyad reported that one of the dyad members was popping bubbles of the wrong colour. From this, we can conclude that none of the dyads exercised a coordination policy in this game – a result we had anticipated, due to the fast pace and competitive nature of the system. Participants frequently collided with each other in this system.

We look in more detail at our Físchlár-DT 1 system for both rules. Figure 5.19 shows the total number of coordination errors that each dyad had for each of the rules of this system. One must note, that since we were missing video footage for Dyad 9, we could not determine how many coordination errors they had. Hence, they were omitted from the graph. From this figure, we can see that there were large variations in the number of coordination errors from dyad to dyad. The blue line indicates the coordination errors for dyads under the "10 Minute" rule. The "Find 10" rule showed much smaller numbers for dyad coordination errors. However, this is most likely due to the fact that dyads generally completed this task in less than twenty minutes (two topics) and hence had less opportunity to incur coordination errors.

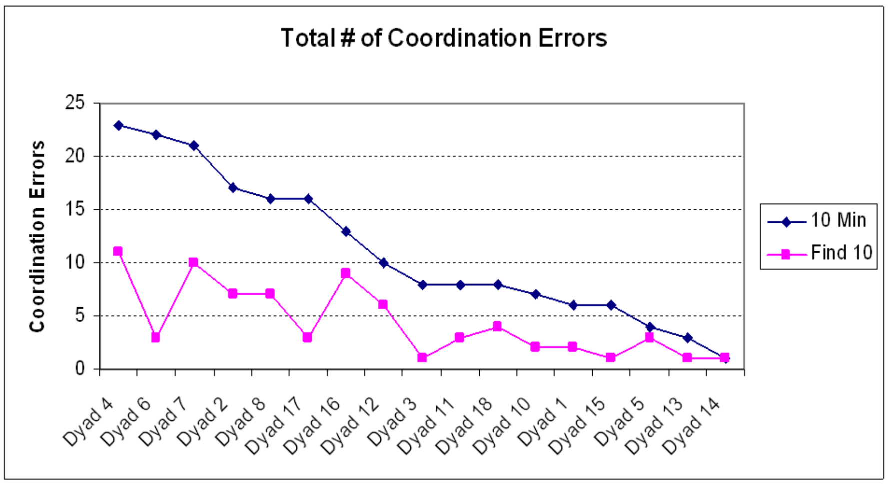


Figure 5.19: *Total coordination errors for dyads on Físchlár-DT 1 (both rules)*

To remove this potential error in our analysis, we normalised the data so that we could compare the coordination errors of the two systems. We used coordination errors per minute as our base time and then multplied all of the resulting numbers by 10 (since the results were less than 1 for most dyads). Figure 5.20 shows the normalised number of coordination errors for each dyad and for each rule. We can see that this resulted in less variation between rules for each dyad than the previous totals graph. However, there was still a great variation in coordination errors between dyads. We can see that Dyad 4 had the highest number of coordination errors in both the "10 minute" and "Find 10" rules. This was followed closely by Dyads 16 and 7 for the "Find 10" rule and Dyads 6 and 7 in the "10 Minute" rule.

Looking at the dyads post-experiment questionnaires for Físchlár-DT 1, we discovered that eleven dyads had similar within-dyad responses when asked

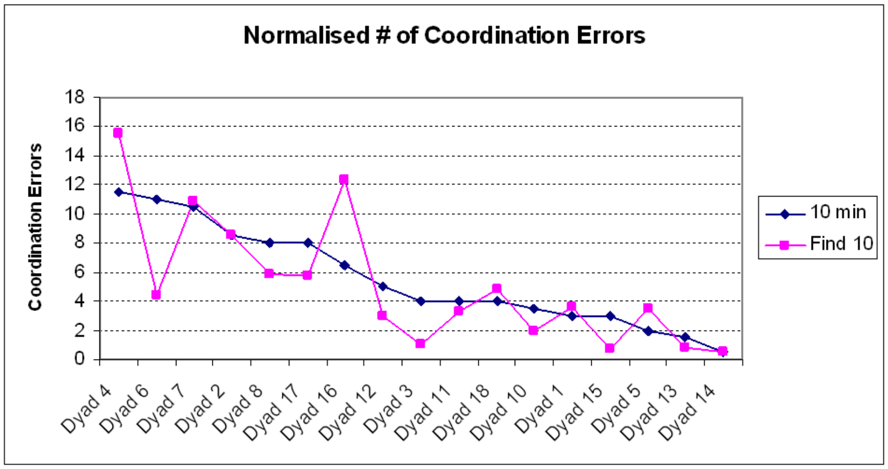


Figure 5.20: *Normalised coordination errors for dyads on Físchlár-DT 1 (both rules)*

about their coordination i.e. either they had no problems with their coordination, or they noticed some problems in their coordination. Dyad 6 User 1 failed to respond to this question and the remaining members of 6 dyads reported very different opinions of their coordination.

In the case of Dyad 4, User 1 reported just a slowing down of the system "When a new search was started by my partner, my interface slowed down". User 2 reported no coordination errors at all, though looking back over the video footage, we saw that User 2 caused many of the coordination errors by executing searches without alerting her partner, resulting in his work being interrupted. This inconsistency between the user's report and the recorded CCTV footage was surprising. Looking at her pre-experiment and personality questionnaire responses and her personality data, we learned that she was in the same age-grouping as her partner (i.e. 20-24), she was highly extravert (82%), highly agreeable (65%), fairly highly conscientious (56%) and scored lowly on the *Openness to Experience* and *Neuroticism* traits (i.e. 34% and

6%). She also reported that she liked working in group situations. Hence, since she was sociable, agreeable and fairly conscientious, this observation seemed confusing. However, upon further analysis of her pre-questionnaire responses, we learned that her course-work did not involve working in groups often, which may have been a factor in her poor coordination skills. Her partner was also highly *Extraverted* (69%), though he scored lowly on the *Agreeableness* and *Conscientiousness* traits (37% and 35% respectively). He communicated a lot and appeared to have good ideas on what to search for. These two users were also in a relationship together, which could have been a factor. Dyad 6 User 2 reported that they "Got better as we went along. Was good to try to establish roles - Pass shots to other person etc.". This was confirmed in their drop in coordination errors from the "10 Minute" rule to the "Find 10" rule.

Dyad 7 acknowledged that they interrupted each other's video playback, but aside from that they thought that their coordination was fine. Dyad 16 did not mention any problems in coordination in their responses, although they scored above average coordination errors for both rules. Those who had particularly low instances of coordination errors (Dyads 3, 13, 14 and 15) mainly reported good coordination, although Dyad 15 User 1 stated "We sometimes got in each other's way. It was difficult to coordinate our efforts".

Figure 5.21 illustrates the number of coordination errors each dyad had on both the Extravert and Introvert interfaces. We can see that Dyads 4 and 17 have the highest number of coordination errors, in particular on the Introvert interface. We had expected that there would be more coordination errors on the Introvert interface, as participants would be less aware of what their partners were doing, since they had their own functions on their own side of the table.

However, we can see that not all dyads had more coordination errors on the Introvert interface in comparison to the Extravert interface. In fact, just eight dyads had more coordination errors on the Introvert interface. Seven dyads had more coordination errors on the Extravert interface. Video footage for Dyad 11's interaction on the Introvert interface was missing, so we put this at zero for the purposes of the graph. Video footage for Dyads 1 and 13 were missing for both interfaces and hence, were omitted completely from the graph.

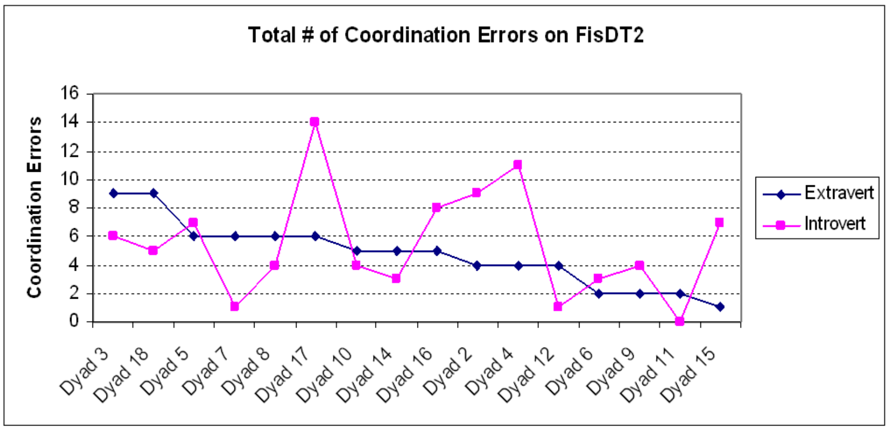


Figure 5.21: *Total coordination errors for dyads on Físchlár-DT 2 (both interfaces)*

The responses to our post-experiment questionnaire contradicted this graph somewhat in the case of both of the dyads who had the highest number of collaboration errors. Just User 2 from Dyad 4 responded and said that coordinating actions was much more difficult on the Extravert interface than the Introvert interface: "Very difficult! Difficult to reach over when we have to share the "Play" button. I much preferred the separate functions". This was surprising since they had a greater number of coordination errors on the Introvert interface as opposed to the Extravert interface. Once again, we saw that User 2 was actually the participant that was causing most of the coordination errors by executing functions without warning her partner, resulting in his work being interrupted.

Both users from Dyad 17 reported that they did not have any problems in collaborating their actions. User 1 reported that: "With separate functions, it was less cluttered around the buttons, and it was easier with the Save Area in the centre". Again, this was surprising since this dyad had a far greater number of interaction errors on the Introvert interface than on the Extravert interface.

The responses from those who scored the lowest number of coordination errors (Dyads 6, 9, 12 and 15,) supported the graphical representation of their coordination. Dyad 6 User 2 reported better coordination on the Extravert interface than on the Introvert interface, which is evident from the graph. Dyads 9 and 12 reported generally good coordination and Dyad 15 User 2 said that their coordination was better on the Extravert interface than on the Introvert interface (due to a lack of awareness on the Introvert system).

Next, we looked at the coordination errors of the dyads to see if they were significantly correlated to each of the personality traits of the "Big Five". We see from Table 5.23 that, using Spearman's rank correlation, only one personality trait was important in one of the systems i.e. increasing coordination errors were significantly correlated to decreasing *Least Agreeable* dyad member score on the Físchlár-DT 2 (Extravert interface). This means that dyads, where both members had a high score on the Agreeableness trait, had less coordination errors than dyads who had at least one very disagreeable member. This would appear to be a logical finding.

	Related Traits	**r_s value**
Fís-DT 1 (10 min)	—	—
Fís-DT 1 (Find 10)	—	—
Fís-DT 2 (Int)	—	—
Fís-DT 2 (Ext)	*Least Agreeable*	-0.59

Table 5.23: *Traits significantly related to coordination errors for each system and associated combination metric.*

Since only one personality trait was significantly correlated to just one of our systems, we can say that, in general, the coordination of dyads is not significantly related to personality.

We chose to look at the age of the participants, to see if this had an impact on the number of coordination errors for dyads on the Fschlr DT 1 and Fschlr DT constraint and interface variants. We chose not to look at gender, since we had only 3 female participants and so meaningful results could not be inferred. The majority of users also had equivalent experience in using the DiamondTouch, with 3 users having used it twice more than the rest of the participants.

	< 20	**20 - 24**	**25 - 29**	**Mixed Age Group**
Fís-DT 1 (10 min)	6	5	10	4
Standard Deviations	3.5	3.4	1.8	3.1
Fís-DT 1 (Find 10)	5	6	6	3
Standard Deviations	3.9	6.2	2.9	2.2
Fís-DT 2 (Int)	2.5	7.33	6	7
Standard Deviations	1.7	2.1	4.2	6.1
Fís-DT 2 (Ext)	4	5.67	3	5.3
Standard Deviations	2	3.1	1.4	0.6

Table 5.24: *Average number of coordination errors per dyad age group and associated standard deviations.*

We see from Table 5.24, that for Físchlár-DT 1 (10 minute rule) the age-grouping 25-29 year olds have the highest average number of coordination errors, with a standard deviation of 1.8. The other age groupings have a similar number i.e. 6, 5 and 4. However, we must note that there were only two dyads whose members collectively fell into the 25-29 year old age grouping. The average number of coordination errors for Físchlár-DT 1 (Find 10 rule) was similar throughout all age-groupings, although the "Mixed Age" group yielded a lower number of coordination errors than the rest. However, we also note that the standard deviations varied greatly between the groups.

For Físchlár-DT 2(Introvert Interface), we see that dyads whose members collectively fell into the "Younger than 20" age-grouping had a far lower average number of coordination errors than the other three. The standard deviation was also quite small, which indicates that this may be significant. However, we also note that the standard deviation for the "Mixed Age" grouping is very high, which indicates that the average scored here may not be a true picture for this group.

Finally, there are no great differences between average coordination errors on the Físchlár-DT 2(Introvert Interface). Again age-group "25-29" year olds had the smallest average, though again, this only constitutes two dyads.

Hence, from this we can say that it would appear that the "Younger than 20" age-group have a lower average number of coordination errors than the other three groups on the Físchlár-DT 2(Introvert Interface).

Our conclusion here is that dyads coordinated their actions better on some interfaces than on others. However, generally their coordination policy was not related to their personality across all tasks.

Q 13. Do the same territoriality tendencies exist regardless of the task or are there cases of some tasks where territoriality is irrelevant in both our competitive and collaborative tasks ?

This question extends previous research that has been conducted into the issue of territoriality on a tabletop interface (see Chapter 2, also see Scott et al., 2004). In previous studies, territoriality tendencies have generally been studied in the case of collaborative tasks. Here, we wished to see if similar territoriality tendencies existed with respect to competitive tasks. Our expectation here was that private and public territories would be more prevalent in collaborative tasks than in competitive tasks.

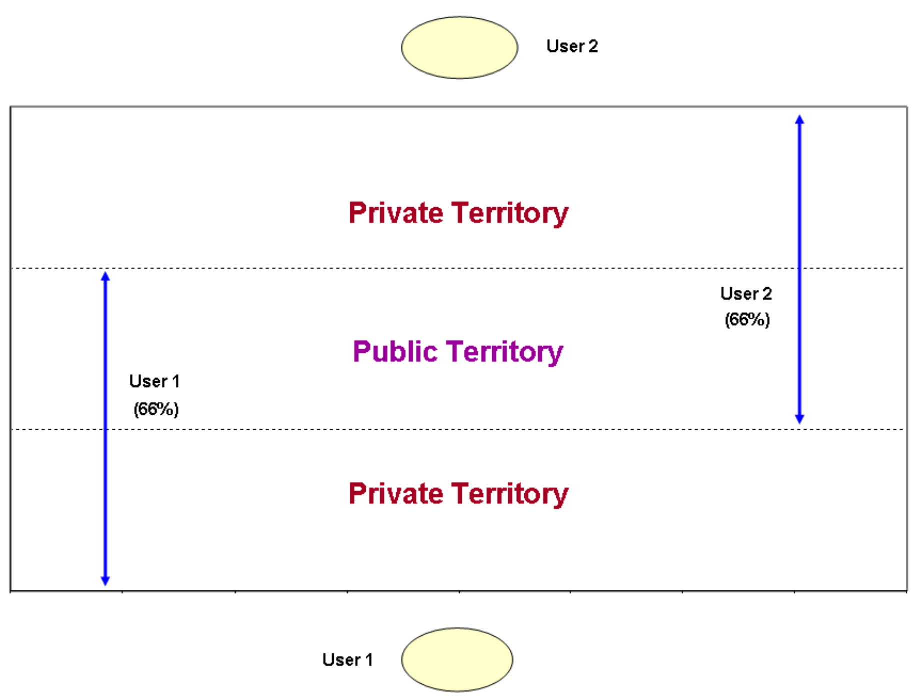


Figure 5.22: *Public and private territories on the DiamondTouch*

To determine if this was the case, we examined the touch-points of all users for all five systems. We looked at the average number of user touch-points at 50% and 66% of the tabletop surface. The reason that we chose 50% of the tabletop was to see if users felt that half the tabletop interface "belonged" to them and the other half "belonged" to their partner i.e. whether they generally interacted with the half of the tabletop closest to them.

We also chose to examine touch-points at 66% of the tabletop surface (from either edge) as we felt that if public and private territories did exist, the edges of the table in front of each user would constitute their private territory (approximately 33% each at either side) and that the central area of the tabletop would make up a public territory (again approximately 33% of the tabletop surface). We felt that if such territories existed, most, if not all of each users' touch-points would be within the 2/3 of the tabletop's surface that was closest to them. Figure 5.22 illustrates this notion of public and private territories in our experimental setting.

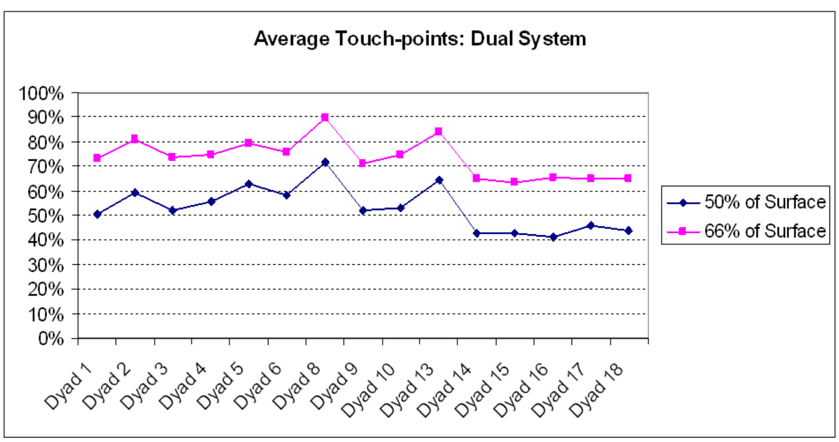


Figure 5.23: *Average touch-points for Pop-A-Bubble Dual Bar interface*

To start, we looked at the touch-points recorded for our **competitive systems** - those being the Pop-A-Bubble game and Competitive Memory Game. To recap, Pop-A-Bubble requires users to "pop" bubbles of their assigned colour as they appear. The bubbles appear randomly all over the tabletop's surface and so people are required to forget about any inhibitions they may have about touching the tabletop surface in the area directly in front of their opponent.

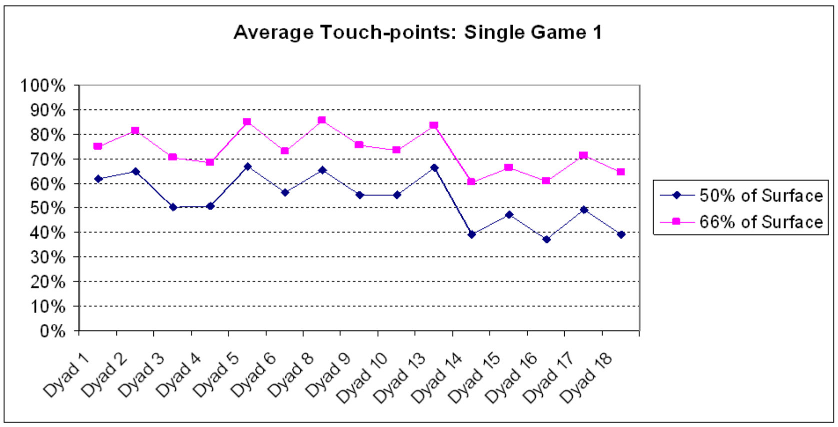


Figure 5.24: *Average touch-points for Pop-A-Bubble Single Bar interface (before swap)*

Hence, we had expected prior to analysis that users would not exhibit a strong notion of private and public territories in this game if they wanted to win. The figures in Appendix C illustrate that this is the case. At the outset, one will notice that we have omitted Dyads 7, 11 and 12 from (reference table) due to the fact that each of these dyads had one colour-blind member. Hence, it would be unfair to examine their touch-points, as they were highly disadvantaged in the game.

One can see that the overall average percentage of touch-points recorded on each user's half of the tabletop was 53% for the dual track-bar system and 54% for both the single track-bar system, with standard deviations of 9% for the dual system and 10% for both single track-bar systems. This tells us that users generally seemed to spend as much time on their opponents half of the table popping bubbles as they did on their own half.

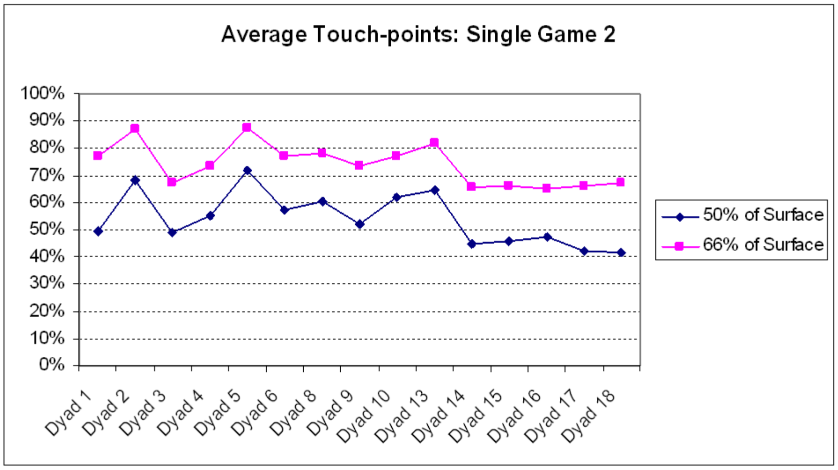


Figure 5.25: *Average touch-points for Pop-A-Bubble Single Bar interface (after swap)*

For the two-thirds of the tabletop closest to each user, we see that the average number of touch-points for the dual and single track-bar systems were 73%, 73% and 74%, with standard deviations of 8% for all 3 variants respectively. However, we must note that in the post-experiment questionnaire responses, five of the participants reported that they found it difficult to stretch far over to the other side of the table or that the table height made it difficult to stretch (Toney and Thomas, 2006).

In a competitive game such as this, where time is of the essence, this could hinder a player's performance and may in fact have made the percentages of touch-points at the 66% section higher than they would have been had the opposite side of the table had been more easily accessed. Another reason for a slightly higher average percentage of touch-points in this region could be that the system did not generate as many bubbles of the user's colour at their opponent's side of the table.

Figures 5.23, 5.24 and 5.25 illustrate these results for each dyad on each

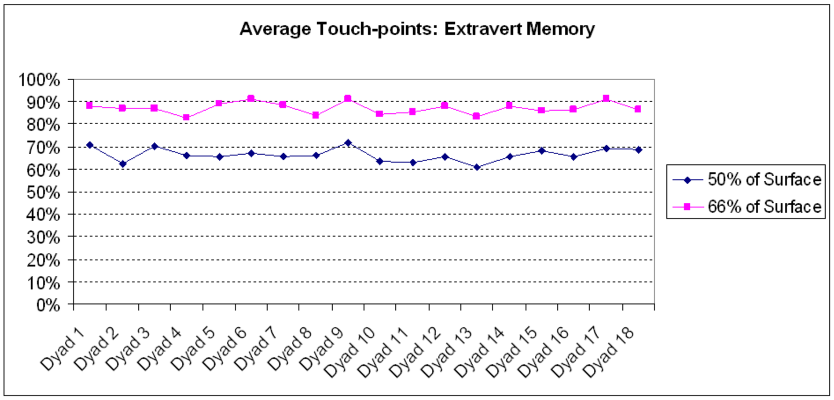


Figure 5.26: *Average touch-points for Competitive Memory (Extravert interface)*

interface variant of the Pop-A-Bubble game.

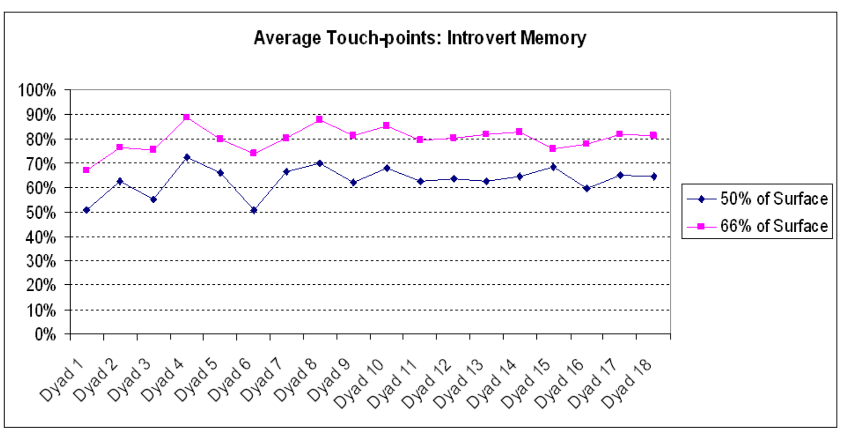


Figure 5.27: *Average touch-points for Competitive Memory (Introvert interface)*

Next we took our other competitive system Competitive Memory game with its two versions - Extravert Memory game and Introvert Memory game. The average percentage of touch-points recorded for dyads on their half of the table-top's surface were 66% for Extravert interface and 63% for Introvert Memory Game with standard deviations of 3% and 6% respectively. For the two-thirds of the tabletop that was closest to each user, the average percentage of touch-points recorded for both Extravert and Introvert Memory Game systems were 87% and 80%, with standard deviations of 3% and 5% respectively.

These figures were much higher than those observed for the Pop-A-Bubble system variants. There are a number of possible reasons for this. Firstly, the random and speedy appearance of bubbles of each user's assigned colour in this highly competitive game, led to them forgetting all about their inhibitions and thoughts about entering their opponents territories. In contrast, the static presentation of identical cards on the tabletop may have led to users preferring to select cards that were closer to them. Also, since some users reported discomfort in stretching, they may have preferred selecting cards that were closer to them.

Figures 5.26 and 5.27 display the average percentages of touch-points for each dyad on these interfaces in each portion of the DiamondTouch surface.

We then examined our **collaborative systems** - those being Collaborative Memory game (both Accuracy and Speed versions), Físchlár-DT 1 and Físchlár-DT 2 (both Introvert and Extravert versions). In Accuracy Memory, the average number of touch-points for all users and all games for their half of the tabletop's surface was 69%, with a standard deviation of 7%.

Speed Memory also saw an average of 70% of all users' touch-points recorded on their half of the tabletop. 86% and 88% of dyads' touch-points were recorded for the two-thirds of the tabletop closest to them on the Accuracy and Speed interfaces, with standard deviations of 3% and 4% respectively.

The average percentages of touch-points for each dyad for each of these rules

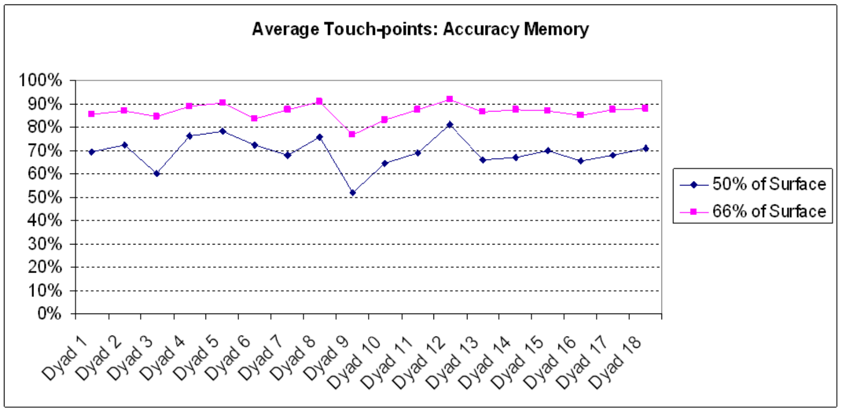


Figure 5.28: *Average touch-points for Accuracy Memory rule*

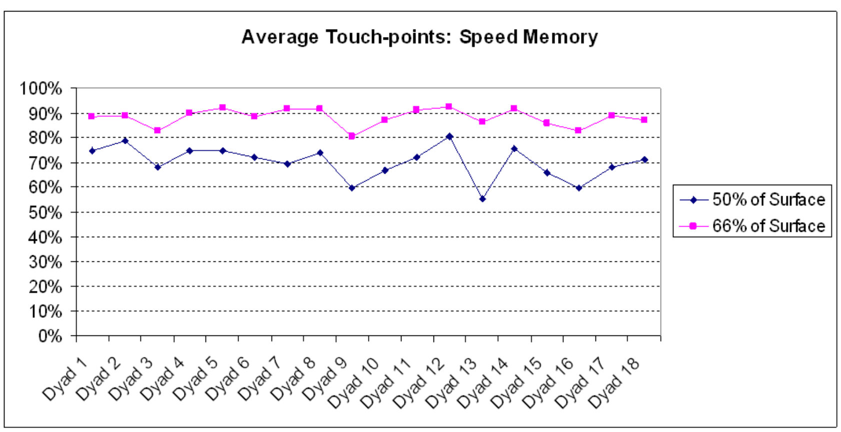


Figure 5.29: *Average touch-points for Speed Memory rule*

and in each territory on the DiamondTouch surface can be seen in Figures 5.28 and 5.29.

For Físchlár-DT 1, an average of 76% of dyads' touch-points were recorded for each user's half of the tabletop, with a standard deviation of 9%. An average of 88% of touch-points for all dyads (with a standard deviation of 4%) were in each user's 66% of the system interface. The breakdown of average touch-points for each dyad for this system can be seen in Figure 5.30.

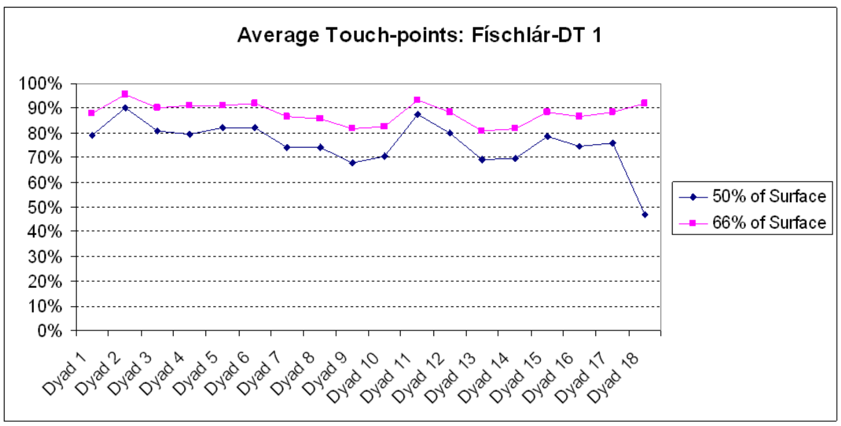


Figure 5.30: *Average touch-points for Físchlár-DT 1*

Físchlár-DT 2's Introvert and Extravert interface variants presented us with a much higher average number of touch-points in each participant's half of the tabletop than Físchlár-DT 1 - these figures being of 89% and 79% with standard deviations of 4 % and 8% respectively. This was supported by a response given by Dyad 9 User 2 when asked about his perception of the coordination between him and his partner - "We didn't get in each other's way as we were using half of the table each".

More evidence of territoriality issues can be seen in the data for the two-thirds of the tabletop closest to each user, with a large proportion of touch-points being recorded – an average of 94% for the Introvert interface and 89% for the Extravert interface with standard deviations of 2% and 5% respectively.

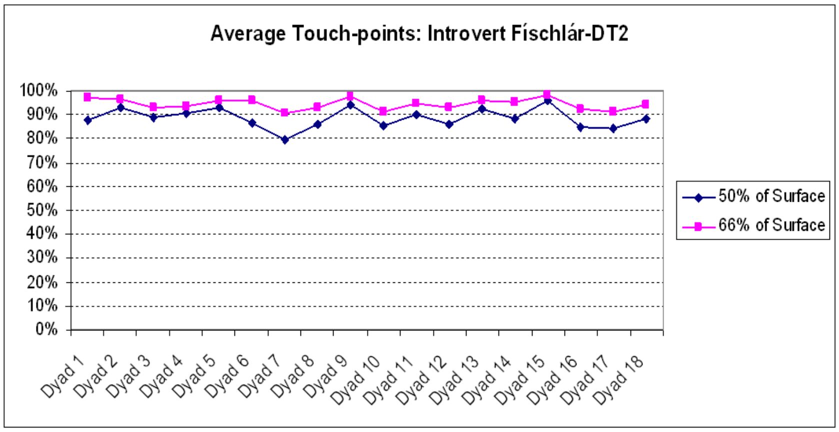


Figure 5.31: *Average touch-points for Físchlár-DT 2 (Introvert interface)*

One of the primary reasons for more user-territoriality in the Introvert interface was that each user had their own set of functions on their own side of the tabletop, whereas the Extravert interface required users to share the functions at either side of the tabletop.

The average number of touch-points per dyad on these two interfaces can be seen in Figures 5.31 and 5.32.

Overall, for competitive systems an average of 59% and 78% of all users' touch-points were noted at the 50% and 66% of the surface closest to them respectively (with standard deviations of 5% of each proportion of touch-points). This indicates that territoriality is only mildly present in competitive systems. Collaborative systems on the other hand displayed much greater evidence of the presence of territoriality with average proportions of 76% and 89% of touch-points (with standard deviations of 5% and 3%) being recorded for 50% and 66% of the tabletop's surface.

This has a direct impact on the design of multi-user, tabletop interfaces. For

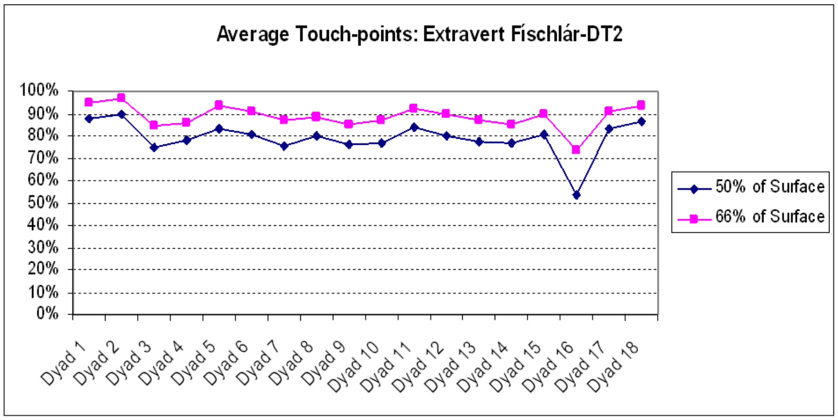


Figure 5.32: *Average touch-points for Físchlár-DT 2 (Extravert interface)*

competitive systems like the Pop-A-Bubble and Competitive Memory game above, territoriality is not hugely important. Hence, consideration for user territories does not affect the placement of objects on such interfaces. However, with regards to our collaborative interfaces, territoriality is much more evident and observed. Hence, the placement of objects such as function areas and widgets should be placed according to the type of collaboration that the designer would like e.g. if the designer wants to cater for and observe user territories in their designs, then widgets and functions that each user needs should be placed in their territories. Likewise, if a designer wants to enable widgets to be shared, then they should place these in public space areas at the centre of the table (see also Scott et al., 2004, and Ringel Morris et al., 2006).

Our conclusion from this is that the notion of territoriality is certainly more evident in our collaborative tasks than in our competitive tasks.

Q 14. Do dyads with certain personalities employ different territoriality techniques than others when performing all of our tasks ?

We studied whether each of the "Big Five" (Johnson, 2008) personality traits of each dyad member, as well as the personality congruence of each dyad, were related to the territoriality tendencies shown by the dyads. We saw from Question 15 above, that the majority of participants exercised territoriality to varying degrees across all systems, and that participant territoriality was more apparent in collaborative systems than in competitive systems. We wished to determine whether the degree to which these territoriality tendencies existed was related to the personalities of each of the participants.

Looking firstly at the territoriality tendencies of individual dyad members, we conducted Spearman's rank correlation method between dyad members' territoriality at 50% of the tabletop and their scores along each of the "Big Five" personality traits for each of the systems. We repeated this for territoriality at 66% of the tabletop. Our sample size for the Pop-A-Bubble systems was 30, since we excluded the dyads containing colour-blind members, who were significantly disadvantaged in the game (Dyads 7, 11 and 12). For the remainder of our systems, our sample size was 36.

Since our sample size was greater than 30 for the remaining systems, we looked at the values of the t-distribution statistics, the critical value of which was 2.042 at $\alpha = 0.05$ (two-tailed test). For Pop-A-Bubble, we looked at the Spearman's rank correlation critical value for our sample size of 30, which was 0.364 for $\alpha = 0.05$ (two-tailed test). Looking at the resulting correlation coefficients and t-statistics that we calculated, we found only one significant correlation at 50% of the tabletop – a positive correlation between increasing percentage territoriality and increasing score along the *Extraversion* trait on the Speed Memory Game (a resulting t-stat value of 2.58).

This was surprising, since we expected that an increase in the percentage of *Extraversion* would result in a higher level of confidence in the individual to use the other half of the tabletop's surface and hence they would exercise less territoriality. However, we recall from Question 2 above, that personality had no significant effect on performance in this system, which was likely as a result of the nature of the system. This may be a reason why we obtained a result that was not intuitive here.

At 66% of the tabletop's surface, we found only two significant correlations between personality traits and territoriality, again on the Speed Memory Game. These were a positive, significant correlation between increasing scores on the *Openness to Experience* personality trait (t-stat value of 2.16) and increasing territoriality and a negative correlation between increasing territoriality and decreasing *Neuroticism* (a t-stat value of -2.07).

We then examined whether correlations existed between the personality congruence of each dyad and their average territoriality at 50% and 66% of the tabletop. We recall that personality congruence was the total of the absolute differences in the scores of each dyad member and their partner along each of the "Big Five" personality traits (see Question 3 above for more details). Again, our sample size was different for the Pop-A-Bubble variations – here our sample size was 15, with a Spearman's rank correlation critical value of 0.545 at α = 0.05 (two-sided test). For the remaining systems and their respective variations, our sample size was 18, with a critical value (i.e. r_s) of 0.476 at $\alpha = 0.05$ (two-tailed test). However, after conducting these Spearman rank correlations, we did not discover any significant relationships between personality congruence and territoriality at either 50% or 66% of the tabletop surface.

To conclude, we have discovered that territoriality solely correlated to the combined *Extraversion* of dyad members only on the Speed Memory Game at 50%. Combined *Openness to Experience* and *Neuroticism* correlated to territoriality at 66% of the tabletop's surface, once again for the Speed Memory system. We found no significant relationships between the personality congruence of the dyad and their territoriality tendencies.

Q 15. Does performance of dyads vary to a greater or lesser extent across the different collaborative tasks used ?

The reason for asking this question is to determine whether dyads performed consistently well/badly across all tasks, relative to each other. In order to answer this question, we look at the performance ranks of the dyads on each of our collaborative tasks. We do not look at performance ranks on competitive systems, since users play against each other on these systems and so we cannot examine a "dyad performance". By placing the dyads in order of decreasing performance into a table for each collaborative system, we can visually see if trends emerge e.g. do some dyads generally perform better on certain types of tasks than others?

Looking at Table 5.25, we can see that the performance ranks of dyads vary across the two types of task - those types being card-matching (Accuracy Memory and Speed Memory) and video searching (the Físchlár-DT systems).

To statistically analyse the variation in dyad performance across systems, we carried out a series of Spearman rank correlations for all performance rank combinations of each interface. The resulting r_s values for these rank combinations are provided in Table 5.26.

Rank	**Accuracy Memory**	**Speed Memory**	**Físchlár-DT 1 (10 min)**	**Físchlár-DT 1 (Find 10)**	**Físchlár-DT 2 (Extravert)**	**Físchlár-DT 2 (Introvert)**
1st	Dyad 17	Dyad 10	Dyad 18	Dyad 17	Dyad 5	Dyad 18
2nd	Dyad 1	Dyad 7	Dyad 6	Dyad 1	Dyad 7	Dyad 13
3rd	Dyad 15	Dyad 1	Dyad 2	Dyad 6	Dyad 6	Dyad 17
4th	Dyad 8	Dyad 12	Dyad 3	Dyad 18	Dyad 17	Dyad 5
5th	Dyad 9	Dyad 14	Dyad 4	Dyad 4	Dyad 2	Dyad 3
6th	Dyad 6	Dyad 15	Dyad 10	Dyad 16	Dyad 18	Dyad 6
7th	Dyad 12	Dyad 16	Dyad 16	Dyad 11	Dyad 4	Dyad 15
8th	Dyad 10	Dyad 8	Dyad 5	Dyad 7	Dyad 16	Dyad 2
9th	Dyad 7	Dyad 18	Dyad 1	Dyad 2	Dyad 10	Dyad 7
10th	Dyad 14	Dyad 6	Dyad 11	Dyad 10	Dyad 8	Dyad 10
11th	Dyad 18	Dyad 17	Dyad 7	Dyad 3	Dyad 11	Dyad 12
12th	Dyad 4	Dyad 3	Dyad 13	Dyad 5	Dyad 12	Dyad 1
13th	Dyad 11	Dyad 11	Dyad 15	Dyad 9	Dyad 3	Dyad 11
14th	Dyad 16	Dyad 5	Dyad 8	Dyad 13	Dyad 14	Dyad 4
15th	Dyad 2	Dyad 13	Dyad 9	Dyad 8	Dyad 1	Dyad 8
16th	Dyad 3	Dyad 4	Dyad 17	Dyad 15	Dyad 13	Dyad 14
17th	Dyad 5	Dyad 2	Dyad 12	Dyad 14	Dyad 15	Dyad 9
18th	Dyad 13	Dyad 9	Dyad 14	Dyad 12	Dyad 9	Dyad 16

Table 5.25: *Performance orderings of dyads on Collaborative Systems*

	Speed Memory	**Fís-DT 1 (10 min)**	**Fís-DT 1 (Find 10)**	**Fís-DT 2 (Extravert)**	**Fís-DT 2 (Introvert)**
Accuracy Memory	0.43	0.44	0.12	0.18	0.19
Speed Memory	x	0.22	-0.06	-0.01	0.12
Fís-DT 1 (10 min)	x	x	-0.55	0.44	0.29
Fís-DT 1 (Find 10)	x	x	x	-0.51	-0.19
Fís-DT 2 (Extravert)	x	x	x	x	0.32
Average r_s					0.14

Table 5.26: *Dyad performance rank correlations across Collaborative Systems*

Our critical value here was 0.476 for $\alpha = 0.05$ and n = 18 (two-tailed test). Hence, we can see from this table, that there were only significant negative correlations between the performance ranks on the Físchlár-DT 1 (10 Minute) system and the Físchlár-DT 1 (Find 10) systems (i.e. dyads who performed well on the “Find 10” rule performed worse on the “10 Minute” rule), as well as a negative correlation between the performance ranks on the Físchlár-DT 1 (Find 10) system and Físchlár-DT 2 (Extravert interface). Three other correlations approached the critical value (i.e. between Accuracy Memory and Speed Memory, Accuracy Memory and Físchlár-DT 1 (10 Min) and Físchlár-DT 1 (10 Min) and Físchlár-DT 2 (Extravert interface)), indicating that dyads performed well across both these types of systems. However, this was not statistically supported.

We then calculated an overall average r_s across all systems, resulting in a value of 0.14. This is a statistically insignificant figure, indicating a large amount of variability in dyad performance across all of the systems and their respective interfaces or rules (i.e. dyad performance rank is not consistent across all task – they will perform better in some tasks than others).

Hence, we can conclude by saying that the performance of dyads varies significantly across the different tasks.

Q 16. How much variability is there in the interaction among dyads across the different collaborative tasks used ?

The purpose of this question is to determine whether dyads interacted with each other consistently highly/lowly across all tasks, relative to each other. We approach this question in the same manner as Question 15 above. However, we look at the ranks of dyad interaction instances and touch-points across our collaborative systems rather than performance ranks. Taking firstly **interaction instances**, we arrange our dyads in order of increasing number of interaction instances (interaction instances per minute in the case of the Accuracy Memory, Speed Memory and Físchlár-DT 1 (Find 10) systems, where the length of the task varied), as in Table 5.30.

Rank	**Accuracy Memory**	**Speed Memory**	**Físchlár-DT 1 (10 min)**	**Físchlár-DT 1 (Find 10)**	**Físchlár-DT 2 (Extravert)**	**Físchlár-DT 2 (Introvert)**
1st	Dyad 13	Dyad 16	Dyad 16	Dyad 16	Dyad 16	Dyad 5
2nd	Dyad 16	Dyad 13	Dyad 5	Dyad 5	Dyad 5	Dyad 3
3rd	Dyad 3	Dyad 14	Dyad 2	Dyad 14	Dyad 3	Dyad 16
4th	Dyad 11	Dyad 17	Dyad 14	Dyad 3	Dyad 7	Dyad 17
5th	Dyad 17	Dyad 9	Dyad 7	Dyad 7	Dyad 17	Dyad 7
6th	Dyad 14	Dyad 3	Dyad 15	Dyad 8	Dyad 15	Dyad 2
7th	Dyad 15	Dyad 7	Dyad 6	Dyad 6	Dyad 13	Dyad 15
8th	Dyad 8	Dyad 11	Dyad 3	Dyad 15	Dyad 2	Dyad 9
9th	Dyad 18	Dyad 10	Dyad 8	Dyad 2	Dyad 9	Dyad 6
10th	Dyad 6	Dyad 18	Dyad 17	Dyad 11	Dyad 11	Dyad 13
11th	Dyad 9	Dyad 8	Dyad 13	Dyad 13	Dyad 8	Dyad 12
12th	Dyad 7	Dyad 15	Dyad 11	Dyad 12	Dyad 12	Dyad 4
13th	Dyad 2	Dyad 1	Dyad 1	Dyad 17	Dyad 6	Dyad 10
14th	Dyad 12	Dyad 6	Dyad 12	Dyad 4	Dyad 4	Dyad 8
15th	Dyad 10	Dyad 12	Dyad 4	Dyad 1	Dyad 10	Dyad 18
16th	Dyad 4	Dyad 2	Dyad 18	Dyad 10	Dyad 18	Dyad –
17th	Dyad 1	Dyad 4	Dyad 10	Dyad 18	Dyad –	Dyad –
18th	Dyad –	Dyad –	Dyad –	–	Dyad –	Dyad –

Table 5.27: *Increasing interaction instance orderings of dyads on Collaborative Systems*

	Speed Memory	Fís-DT 1 (10 min)	Fís-DT 1 (Find 10)	Fís-DT 2 (Extravert)	Fís-DT 2 (Introvert)
Accuracy Memory	0.7	0.42	0.48	0.59	0.44
Speed Memory	x	0.3	0.39	0.6	0.43
Fís-DT 1 (10 min)	x	x	0.88	0.8	0.8
Fís-DT 1 (Find 10)	x	x	x	0.78	0.7
Fís-DT 2 (Extravert)	x	x	x	x	0.92
Average r_s					0.60

Table 5.28: *Dyad interaction instance rank correlations across Collaborative Systems*

This provides us with a visual indication as to whether dyads have a consistent number of interaction instances across tasks. We can see from this table, that many dyads retained the same or a similar rank position across the systems e.g. Dyad 16 had the lowest number of interaction instances for four out of the six systems.

Once again, we performed a Spearman rank correlation between the ranks for each system. The resulting correlation coefficients (r_s values) are shown in matrix format in Table 5.28.

Due to the corruption of some of our video recordings, we had different sample sizes for each rank correlation between the systems. The different sample sizes resulted in different critical values for these correlations. Table 5.29 below, displays the critical values for each rank correlation coefficient.

We can see from a comparison of the two tables, that there are strong correlations of dyad interaction instance ranks on the Accuracy Memory game and dyad interaction instance ranks on the Speed Memory and Físchlár-DT (Extravert interface). There is a strong correlation between ranks on the Speed Memory game and Físchlár-DT 2 (Extravert interface). Dyad interaction instance ranks on Físchlár-DT 1 (10 min) were strongly correlated to interaction

	Speed Memory	**Fís-DT 1 (10 min)**	**Fís-DT 1 (Find 10)**	**Fís-DT 2 (Extravert)**	**Fís-DT 2 (Introvert)**
Accuracy Memory	0.509	0.509	0.509	0.545	0.545
Speed Memory	x	0.509	0.509	0.545	0.545
Fís-DT 1 (10 min)	x	x	0.509	0.545	0.545
Fís-DT 1 (Find 10)	x	x	x	0.545	0.545
Fís-DT 2 (Extravert)	x	x	x	x	0.545
Average critical value					0.545

Table 5.29: *Critical values for interaction instance rank correlations across Collaborative Systems*

instance ranks on Físchlár-DT 1 (Find 10), Físchlár-DT 2 (Extravert interface) and Físchlár-DT 2 (Introvert interface), as was well as strong correlations noted between interaction instance ranks on the Físchlár-DT 1 (Find 10) and ranks on Físchlár-DT 2 (Extravert interface) and Físchlár-DT 2 (Introvert interface). Finally, there was a very strong correlation between dyad interaction instance ranks on Físchlár-DT 2 (Extravert interface) and ranks on Físchlár-DT 2 (Introvert interface).

We also calculated our average critical value across all rank combinations. Since the average r_s value (0.60) was greater than the average critical value (0.545), we can say that there was a strong correlation between the interaction instances of dyads across all systems i.e. there was little variability. Hence, dyads who communicate a lot relative to others on one system, will still communicate a lot relative to others on another system.

We now look at our dyads, ordered by increasing number of **touch-points** on each system and respective interface(see Table 5.31). Below this, Table 5.31 displays the r_s values for each combination of these ranks, in the form of a matrix. Since our sample size here is constant (n = 18 for all ranks), our critical value is 0.476, for $\alpha = 0.05$ (two tailed).

Rank	**Accuracy Memory**	**Speed Memory**	**Físchlár-DT 1**	**Físchlár-DT 2 (Extravert)**	**Físchlár-DT 2 (Introvert)**
1^{st}	Dyad 1	Dyad 9	Dyad 10	Dyad 1	Dyad 7
2^{nd}	Dyad 17	Dyad 17	Dyad 12	Dyad 9	Dyad 1
3^{rd}	Dyad 12	Dyad 4	Dyad 1	Dyad 13	Dyad 9
4^{th}	Dyad 3	Dyad 2	Dyad 14	Dyad 11	Dyad 18
5^{th}	Dyad 15	Dyad 6	Dyad 8	Dyad 10	Dyad 8
6^{th}	Dyad 10	Dyad 11	Dyad 18	Dyad 16	Dyad 16
7^{th}	Dyad 5	Dyad 15	Dyad 9	Dyad 4	Dyad 11
8^{th}	Dyad 14	Dyad 13	Dyad 13	Dyad 14	Dyad 4
9^{th}	Dyad 2	Dyad 14	Dyad 7	Dyad 8	Dyad 14
10^{th}	Dyad 6	Dyad 7	Dyad 5	Dyad 12	Dyad 3
11^{th}	Dyad 9	Dyad 5	Dyad 11	Dyad 3	Dyad 10
12^{th}	Dyad 7	Dyad 8	Dyad 16	Dyad 7	Dyad 5
13^{th}	Dyad 8	Dyad 1	Dyad 4	Dyad 6	Dyad 17
14^{th}	Dyad 18	Dyad 16	Dyad 3	Dyad 17	Dyad 13
15^{th}	Dyad 11	Dyad 3	Dyad 17	Dyad 18	Dyad 15
16^{th}	Dyad 4	Dyad 18	Dyad 6	Dyad 2	Dyad 6
17^{th}	Dyad 16	Dyad 12	Dyad 15	Dyad 5	Dyad 12
18^{th}	Dyad 13	Dyad 10	Dyad 2	Dyad 15	Dyad 2

Table 5.30: *Touch-point orderings of dyads on Collaborative Systems*

	Speed Memory	**Fís-DT 1**	**Fís-DT 2 (Extravert)**	**Fís-DT 2 (Introvert)**
Accuracy Memory	-0.19	0.11	-0.24	-0.25
Speed Memory	x	-0.59	-0.08	-0.14
Fís-DT 1	x	x	0.54	0.41
Fís-DT 2 (Extravert)	x	x	x	0.46
Average r_s				0.001

Table 5.31: *Dyad touch-points rank correlations across Collaborative Systems*

From our r_s values in Table 5.31, we can see that significant correlations were found for dyad touch-point ranks on Speed Memory and Físchlár-DT 1, between Físchlár-DT 1 and Físchlár-DT 2 (Extravert interface). However, our average r_s value across all rank combinations is much less than our critical value - in fact, at 0.001 it is close to 0, indicating that their is almost no relationship between the dyad touch-point orderings across systems.

From the above analysis, we can conclude that there is little variability in dyad interaction instances across systems. In contrast, our low average Spearman correlation coefficient indicates that there is high variability in the touch-points of dyads across systems.

Hypothesis Support / Disproval

The answers that we have obtained above contributed to our support or disproval of the hypothesis "Dyads perform different tasks in a different manner and this is related to their personality.". We have seen that the relative performance of dyads was significantly different under different constraints in our Collaborative Memory game. Dyad interaction in the form of touch-points only related to performance on Accuracy Memory, Speed Memory and Físchlár-DT 2 (Introvert Interface), while communication did not statistically affect performance on any of our collaborative tasks.

Coordination was generally unrelated to personality with the exception of Físchlár-DT 2 (Extravert Interface), where *Agreeableness* was significantly correlated to coordination. Some dyads coordinated their actions better on the interfaces to our collaborative search tasks than others. Territoriality was exercised more by dyads on collaborative systems than on competitive systems, while territoriality was only related to personality on our Speed Memory system i.e. at combined *Extraversion* at 50% of the tabletop's surface, as well as *Neuroticism* and *Openness to Experience* at 66% of the tabletop's surface.

From these answers, we can conclude that dyads do perform different tasks in a different manner, though this is generally unrelated to their personality. Hence, we can only support the first half of this hypothesis, but cannot support its broad relationship to dyad personality.

5.6 Chapter Summary

In this chapter, we provided an extensive analysis of the user experiments that we conducted on our five multi-user systems, involving 18 dyads. We introduced our data-set – what was recorded and how this data was used in order to answer our sixteen hypotheses sub-questions, which is turn would enable us to prove or disprove the three main hypotheses that we posited in Chapter 3.

These user-experiments were based on previous research conducted in our three encompassing areas of HCI, Groupware Technology and Personality Psychology as discussed in Chapter 2. We analysed the performance, interaction, personalities and user questionnaires of the 18 dyads that completed tasks on these systems and identified whether correlations existed between these data-sets. Our analysis focussed more on our collaborative tasks, which were concerned with the workings of the dyad collectively, as opposed to our competitive tasks, where work was carried out on an individual basis.

The results of this analysis provided us with answers to our sixteen hypotheses sub-questions. Our main findings from these were that personality traits did impact performance and interaction as shown in hypothesis sub-questions 1, 2, 3, 4. Subsequently, we could fully verify our first hypothesis (*The personality composition of a dyad impacts the performance of that task, or in other words, dyads composed of certain personality types will perform tasks better than others*).

We found that the interface preferences of our dyads did always not reflect their personality along the *Extraversion* personality trait in the cases of Competitive Memory and Físchlár-DT 2 interface variations. However, we did find that when given two interfaces or task constraints, they performed better on the version that they preferred. Dyad interaction data did not generally relate to users' stated opinions on a system, with the exception of the Físchlár-DT 2 Introvert interface, which statistically showed that those who liked the system communicated more. Our Pop-A-Bubble system also displayed a tendency for users to physically interact with the tabletop's surface less frequently on the interface that they preferred. Hence, we could partially verify our second (*Dyads with certain personality types will prefer and work better on certain interfaces*).

Finally, the answers to the sub-questions of our third hypothesis (*Dyads perform different tasks in a different manner and this is related to their personality*) enabled us to partially verify it i.e. dyads did perform our tasks in a different manner. This was given by the fact that more territoriality was observed in the collaborative tasks than in the competitive tasks, though this was only related to the personality traits of our dyads in the Speed Memory Game. Also, some of the dyads coordinated their actions better than others in collaborative search tasks, though this was generally unrelated to their personality traits. Touch-point data significantly correlated to performance on Accuracy Memory, Speed Memory and Físchlár-DT 2 (Introvert interface). Also, dyads received different relative performance when working under two different constraints imposed on a task. In the next chapter, we provide a general conclusion to the thesis and list some potential avenues for future work based on these experiments.

Chapter 6

Conclusions

As we have seen, multi-user computing technologies have presented us with many new design challenges and opportunities for applications and their associated interfaces. Not only must we cater to the needs, preferences and expectations of a single user working on a single application, but with the emergence of these new multi-user technologies, we must consider both the individual and the combined needs, preferences and expectations of groups people working on the same application, together.

To determine what the most important design issues in collaborative group applications are, we must consider the setting in which these technologies are used, for example, are they being used in a collocated, synchronous setting, or a distributed, asynchronous setting? Differences in the environment in which these technologies are used present different challenges that must be overcome in order to provide a successful and satisfactory user experience. For instance, a multi-user technology utilised in a distributed, synchronous setting requires that each user's awareness of the other group members' actions is explicitly provided for and maintained. This is achieved in the design of such a system by providing awareness cues to all users so that their actions can be better coordinated, one person's work does not overwrite or impair the other's and consistency is visually maintained.

In a collocated, synchronous setting, awareness cues are naturally provided, since in this setting, one person can see the actions of another. However, designers can strengthen awareness cues in this setting, for instance, in placing shared buttons and widgets on the application interface. In doing this, users are required to interact with each other more and so user actions are more pronounced.

Awareness is just one example of a multi-user interaction issue that must be considered when designing such applications. Many more exist, including the tendencies of users to exercise territoriality in the execution of a task, the division of labour among group members, the coordination policy that could or must be exercised by the group, the placement of objects, buttons and widgets on the interface and the orientation of these.

In this thesis, we presented a qualitative study conducted on a collocated, synchronous, multi-user tabletop device, called the DiamondTouch. In conducting this study, involving 18 dyads working on different interfaces or task constraints with five systems, our intention was to develop a deeper understanding of the interactions of dyads on this type of groupware technology, their performance on different kinds of tasks and how these two aspects relate to the personalities of the dyad members, both on an individual and collective basis.

Below is a list of 16 summarised conclusions that we drew from the results of our analysis in Chapter 5. These are:

1. *Extraversion* is unrelated to performance or normalised touch-points in collaborative systems, but has an effect in competitive systems. It significantly relates to interaction instances (i.e. communication).

2. *Agreeableness*, *Conscientiousness* and *Openness to Experience* were related to dyad performance on certain systems.

3. Dyads with similar personality types generally do not outperform dyads with differing personality types.
4. Certain personality traits from the remaining four "Big Five" personality traits were more related to touch-points on different systems. Personality was only related to interaction on our Físchlár-DT 2 system interfaces.
5. Dyads develop a similar impression of a system about half the time.
6. Both introvert and extravert users generally liked more aspects of our introvert interfaces (e.g. colours and layout on Competitive Memory).
7. The majority of dyads performed better on the system interface or constraint that they preferred.
8. Interaction data does not generally relate to users' interface or constraint preferences across all our tasks.
9. Imposing different rules on a system task does impact on the performance of a dyad.
10. Dyads have more interaction instances in collaborative systems than in competitive systems.
11. The number of interaction instances recorded generally does not relate to performance in collaborative tasks, though the number of touch-points does relate to performance in some tasks.
12. Dyads coordinated their actions better on some interfaces than on others. Generally coordination was not related to personality.
13. Could not be scientifically answered.
14. The notion of territoriality is more evident in our collaborative tasks than in our competitive tasks.
15. Territoriality was generally unrelated to personality traits (it was just related in Speed Memory) and completely unrelated to personality congruence.

16. The performance of dyads varies significantly across the collaborative systems.

17. There is little variability in dyad interaction instances, but great variability in dyad touch-points across systems.

These 16 individual small conclusions (we omit number 13) need to be aggregated and combined in some way so from these points, we drew a mind-map in an attempt to group our outcomes under general headings. Figure 6.1 below illustrates the mind-map that we constructed. We can see from this diagram, that our outcomes can be grouped under five headings. This provides us with a guide for condensing all 16 of the outcomes listed above into a smaller number of statements, that we can assert as the conclusions of our study and its associated analysis.

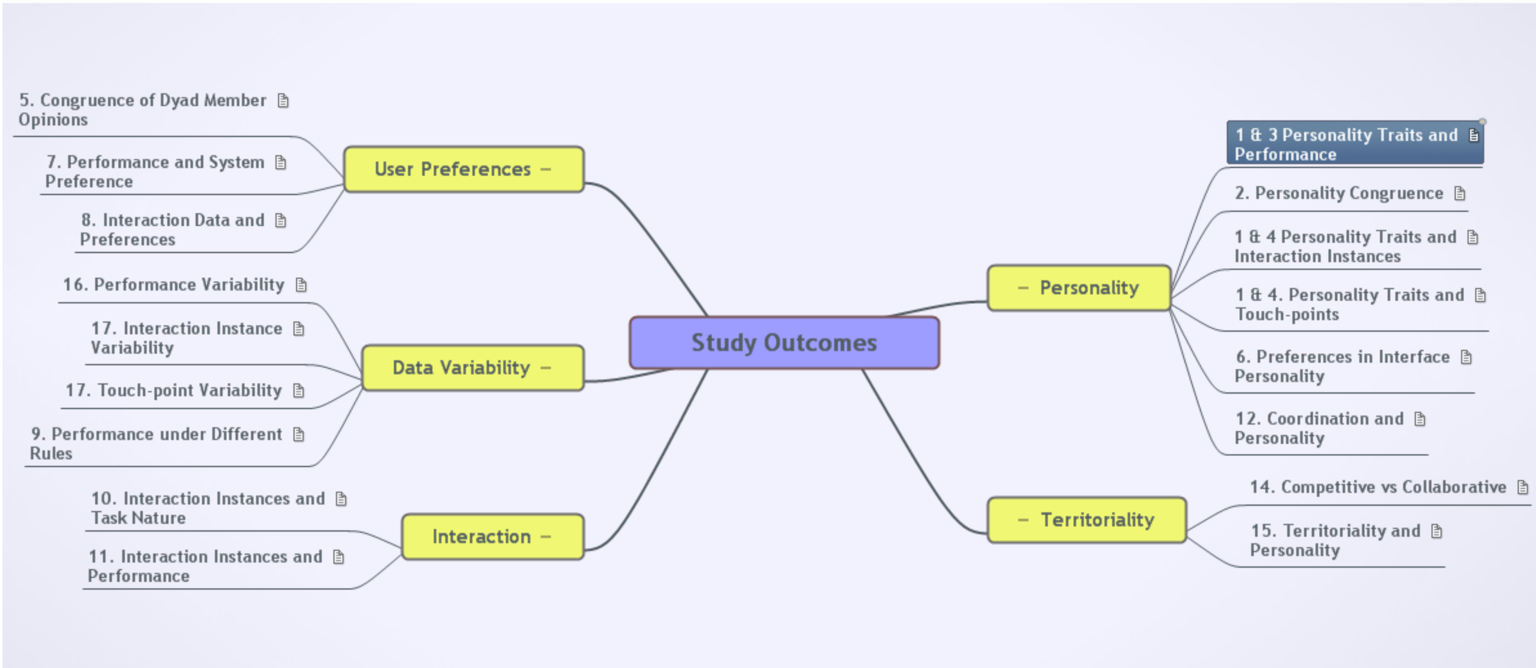


Figure 6.1: *Groupings of our study outcomes*

From looking at our outcomes in this manner, we condensed our overall conclusions into the following six major points, each of which we discuss:

1. Some personality traits affect performance on our collaborative tasks (i.e. *Openness to Experience*, *Conscientiousness* and *Agreeableness*), while others do not (i.e. *Extraversion* and *Neuroticism.*)

We recall from the answer to our sub-hypothesis Question 2, that high levels of *Openness to Experience* and *Conscientiousness* negatively affected performance on our search tasks and that high levels of *Conscientiousness* and *Agreeableness* positively affected performance on Accuracy Memory. The results for our search task were surprising. What was also surprising, was the fact that *Extraversion* had no effect on performance in our collaborative tasks, given the fact that the collaborative tasks were highly social in nature. However, we also saw that the *Extraversion* trait was highly significant in determining the amount of communication between the dyads (i.e. the number of interaction instances between the dyads).

Hence, it would appear that when constructing a dyad that will perform well on a collaborative task, such as a search task, on a tabletop interface, the three most significant traits to look at are their *Openness to Experience*, *Conscientiousness* and *Agreeableness*. This outcome provides an interesting and important contribution to the area of personality psychology.

2. Generally, the similarity of the dyad members' personalities does not affect their performance in our collaborative search and card gaming tasks.

We can say from this that it really does not matter whether dyads are similar in terms of their personality or very diverse – generally, they will not perform better in either case. This was not hugely surprising, since it is possible that the performances of dyads has more to do with their own personal skills, personality traits and educational background, rather than their personality similarity. We note that dyad similarity with respect to their personality scores did have an impact on performance on the Accuracy Memory system, which would lead us to believe that where accuracy is particularly important, having a dyad with a more similar personality is more desirable than having a dyad comprising of very different personality trait scores.

These results in terms of the performance data as related to personality traits provide a meaningful contribution to the area of Personality Psychology. The inferences drawn indicate that personality has an important role to play in predicting the performances of dyads on the tasks we used that didn't impose a timing constraint.

3. Dyads perform better on a system interface that they collectively prefer, but their preference does not affect their level of interaction.

This would appear to be an intuitive conclusion, since if users prefer an interface, they would enjoy working on it more and so would be more productive and effective at the task they are doing. We also noted from our analysis that the majority of dyads prefered interfaces that displayed more introvert qualities. Hence, we can conclude from this that people expressed a stronger preference for interfaces that portrayed more introvert attributes than extravert attributes (e.g. colours used) and that this preference affects their performance. This is something that can certainly be taken into consideration when designing interfaces in order to maximise dyad performance.

In terms of interaction, dyad members generally interact a lot in general, and it is likely that they will continue to interact and communicate regardless of the interface they are working on. We recall that the scores of dyads along the *Extraversion* personality trait had a significant impact on the number of interaction instances annotated for that dyad. Hence, we can say that the level of interaction is more related to their personality than it is to their interface preference.

These findings have an impact on the design of multi-user tabletop interfaces, where dyads with levels of *Extraversion* that directly impact their interaction (as reflected in Tables 5.2 and 5.3) would require interfaces that support this interaction. This would impact the types of widgets used, the display and placement of objects on the interface.

4. The relative amount of interaction across dyads does not change much, but the relative dyad performances and their amount of touch-points does vary across tasks.

This is again an intuitive finding. In terms of the interaction of the dyads, we can see that dyads who generally communicate a lot, will still communicate a lot, relative to other dyads in all tasks i.e. if they have most interaction instances on one system, they will generally have most interaction instances on all other systems, in comparison to the other dyads. It also seems intuitive that dyads would perform differently relative to each other, on different tasks. This would indicate that our choice of systems to use for our experiments were heterogenous in terms of performance, where different dyads performed relatively better in some systems than in others, possibly due to the different skills and personality trait combinations that each dyad collectively had. Also, since we found that touch-points and performance were highly positively correlated, it is unsurprising that the relative amount of touch-points recorded varied greatly across dyads for each system.

5. The amount and type of interaction within dyads is not related to their performance in collaborative tasks and there is actually more

interaction in collaborative rather than in competitive tasks

Based on conclusion 4 above, this would appear to be a logical statement, since relative dyad interaction did not vary across tasks, but performance did. This statement is also reinforced by the fact that *Extraversion* did not affect dyad performance on collaborative tasks, but did affect dyad interaction, indicating that personality, rather than performance, affects dyad interaction. The fact that more dyad interaction was recorded on collaborative systems than on competitive systems is also unsurprising, since communication is an essential element in completing a collaborative task, whereas very little communication occurs when one person is competing against another.

6. Territoriality is more evident in collaborative rather than in competitive systems, and territoriality is generally unrelated to personalities

The first part of this statement is intuitive, since collaborative tasks can be broken into sub-tasks, which each user works on in their own space. Intruding on this space could be seen as being socially inconsiderate or even socially inept. However, since people are only concerned with their own success in a competitive system, there is less of a notion of personal space. The key in a competitive task is to outsmart your opponent, which often takes the form of using objects, in locations that would usually be deemed as the other person's private space. The fact that territoriality is generally unrelated to personalities is a little surprising, as one might expect that, for instance, more extraverted people would be more likely to reach into another person's personal space, due to their sociable nature. However, it would appear that in this case, social protocols and innate behaviours are more important than personality traits with regards to territoriality.

This result directly affects the layout of interfaces (both competitive and collaborative), which should cater to the territoriality of users of such systems. For competitive systems like the Pop-A-Bubble and Competitive Memory game above, territoriality is not hugely important. Hence, consideration for user territories does not affect the placement of objects on such interfaces. However, with regards to our collaborative interfaces, territoriality is much more evident and observed. Hence, the placement of objects such as function areas and widgets should be placed according to the type of collaboration that the designer would like e.g. if the designer wants to cater for and observe user territories in their designs, then widgets and functions that each user needs should be placed in their territories. Likewise, if a designer wants to enable widgets to be shared, then they should place these in public space areas at the centre of the table

We acknowledge that the number of dyads that we selected to participate in this study was relatively small and that generalised, statistically-supported statements on the workings and preferences of all dyads must be made tentatively. However, due to the fact that our experiments were carried out over a significant period of time, using the same people throughout, we feel that this justifies our support for the observations that we made, the relationships between variables that we identified and the conclusions that we have arrived at.

6.1 Hypotheses Revisited

We recall from the early part of this thesis, that our sub-hypotheses questions enabled us to prove or disprove the three hypotheses that we asserted in Chapter 3. We now reiterate these hypotheses and state whether the analysis of our data gathered leads to the support or otherwise of these hypotheses.

Hypothesis 1: The personality composition of a dyad impacts the performance of that task, or in other words, dyads composed of certain personality types will perform tasks better than others.

From our answers to each of the sub-questions of this hypothesis and from our overall conclusion 1 above, we can say dyads with certain personality types do achieve better performance in tasks compared to others, so it is proven. That is, that *Openness to Experience*, *Conscientiousness* and *Agreeableness*), affected performance on our collaborative tasks, while *Extraversion* and *Neuroticism* did not.

Hypothesis 2: Dyads with certain personality types will prefer and work better on certain interfaces.

From looking at the answers to the sub-questions of this hypothesis, we saw that dyads did not intuitively prefer certain interfaces over others. However, we did find that dyads generally did work better on an interface that the members collectively preferred, which is also stated in our overall conclusion 3 above. Hence, we can only support the second part of this hypothesis i.e. that dyads work better on certain interfaces.

Hypothesis 3: Dyads perform different tasks in a different manner and this is related to their personality.

Our analysis led us to the conclusion that dyads do perform different tasks in a different manner, though this is generally unrelated to their personality. Therefore, we can only support the first half of this hypothesis, but cannot support its broad relationship to dyad personality.

6.2 Future Work

The qualitative study that we have presented in this thesis was intended to create a foundation for future research into the workings of dyads on multi-user, collaborative, tabletop technologies, in particular, with regards to the personalities of the dyads. There are many avenues that can be taken in future research in this novel and increasingly popular area. We list some of these below.

1. **Increase the sample size**

   To statistically reinforce the findings of our study here, future studies could be undertaken with a greater sample size.

2. **Increase group size**

   In this thesis, we studied the performances and interaction of dyads and the relationships of these to the personalities of the dyads. Further research in this area could look at the possibility of conducting studies into the workings of groups of three or four people, with this regard.

3. **Look further into the interaction of the dyads**

   Rather than just looking at the overall interaction of dyads as they complete multi-user tasks, a more detailed analysis of the individual interaction instances could be undertaken in relation to the personalities of group members e.g. do more *extravert* people ask more questions while completing a task in comparison to more *introvert* people.

4. **Examine different types of tasks**

   In our study, we looked at simple collaborative and competitive games, as well as more sophisticated variations of a search task. However, there are many more existing and potential tabletop applications that support group-work, some of which have been described in Chapter 2. Studies on these applications to determine the relationship between the personalities of users and their performance and satisfaction on using these systems is certainly a great research opportunity.

5. **Conduct experiments over a longer period of time** While we chose to recruit participants in pairs to encourage people to display their personality as quickly as possible, a longer-term study could be conducted, where people who are unknown to each other are brought together to conduct tasks over a longer period of time.

6. **Conduct experiments using a task-oriented design** We used a system-oriented approach in evaluating our search tasks. Future work could involve looking at a task-oriented approach, or a hybrid of both system and task-oriented approaches.

7. **Look at Dyad Performance, Preference and Interaction from a Management Science Perspective** Here, we looked at the effect that personality psychology had on the performance, preferences and interactions of dyads on our games and search tasks. It would be interesting to see how methods taken from management science would affect these aspects of dyad interaction, performance and preferences would have. It would be particularly interesting to look at the roles each dyad member acquires during the execution of these tasks e.g. the Belbin Team Inventory.

The list above is not exhaustive, but does give us a glimpse at the potential studies that could be undertaken to build on and progress the work done in this thesis.

Appendices

Appendix A

Sample Participant Questionnaire Responses

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Blank.

Appendix B

TRECVid 2005 Search Topics

Topic 0149: Find shots of Condoleeza Rice

Topic 0150: Find shots of Iyad Allawi, the former prime minister of Iraq

Topic 0151: Find shots of Omar Karami, the former prime minister of Lebannon

Topic 0152: Find shots of Hu Jintao, president of the People's Republic of China

Topic 0153: Find shots of Tony Blair

Topic 0154: Find shots of Mahmoud Abbas, also known as Abu Mazen, prime minister of the Palestinian Authority

Topic 0155: Find shots of a graphic map of Iraq, location of Bagdhad marked - not a weather map

Topic 0156: Find shots of tennis players on the court - both players visible at same time

Topic 0157: Find shots of people shaking hands

Topic 0158: Find shots of a helicopter in flight

Topic 0159: Find shots of George W. Bush entering or leaving a vehicle (e.g., car, van, airplane, helicopter, etc) (he and vehicle both visible at the same time)

Topic 0160: Find shots of something (e.g., vehicle, aircraft, building, etc) on fire with flames and smoke visible

Topic 0161: Find shots of people with banners or signs

Topic 0162: Find shots of one or more people entering or leaving a building

Topic 0163: Find shots of a meeting with a large table and more than two people

Topic 0164: Find shots of a ship or boat

Topic 0165: Find shots of basketball players on the court

Topic 0166: Find shots of one or more palm trees

Topic 0167: Find shots of an airplane taking off

Topic 0168: Find shots of a road with one or more cars

Topic 0169: Find shots of one or more tanks or other military vehicles

Topic 0170: Find shots of a tall building (with more than 5 floors above the ground)

Topic 0171: Find shots of a goal being made in a soccer match

Topic 0172: Find shots of an office setting, i.e., one or more desks/tables and one or more computers and one or more people

Appendix C

Users' Personality Profiles

		Openness	Conscient-iousness	Extraver-sion	Agreeable-ness	Neurot-icism
Dyad 1	**User 1**	70%	48%	80%	80%	47%
	User 2	84%	61%	76%	57%	23%
Dyad 2	**User 1**	15%	41%	45%	41%	45%
	User 2	29%	83%	78%	49%	11%
Dyad 3	**User 1**	0%	27%	29%	80%	89%
	User 2	19%	31%	34%	16%	62%
Dyad 4	**User 1**	34%	56%	82%	65%	6%
	User 2	86%	35%	69%	37%	62%
Dyad 5	**User 1**	70%	6%	38%	0%	91%
	User 2	67%	16%	3%	41%	50%
Dyad 6	**User 1**	12%	63%	83%	65%	76%
	User 2	24%	56%	43%	70%	76%
Dyad 7	**User 1**	15%	36%	66%	62%	39%
	User 2	5%	65%	59%	38%	74%
Dyad 8	**User 1**	77%	29%	55%	79%	59%
	User 2	55%	53%	59%	50%	15%
Dyad 9	**User 1**	35%	79%	51%	86%	82%
	User 2	69%	62%	28%	54%	54%
Dyad 10	**User 1**	30%	48%	70%	47%	28%
	User 2	38%	31%	57%	57%	83%
Dyad 11	**User 1**	84%	80%	53%	79%	41%
	User 2	50%	43%	68%	45%	48%
Dyad 12	**User 1**	20%	48%	68%	57%	15%
	User 2	61%	16%	34%	35%	91%
Dyad 13	**User 1**	93%	91%	4%	22%	34%
	User 2	1%	5%	41%	24%	52%
Dyad 14	**User 1**	24%	70%	45%	18%	56%
	User 2	67%	98%	87%	11%	1%
Dyad 15	**User 1**	56%	43%	38%	54%	31%
	User 2	21%	51%	9%	60%	35%
Dyad 16	**User 1**	7%	58%	45%	33%	41%
	User 2	13%	48%	0%	52%	82%
Dyad 17	**User 1**	19%	70%	54%	16%	45%
	User 2	0%	68%	57%	43%	50%
Dyad 18	**User 1**	1%	2%	43%	38%	80%
	User 2	17%	77%	92%	68%	14%

Table C.1: Results of Personality Profiling

Appendix D

Dyad Territoriality on DiamondTouch

	50 % of Tabletop			66 % of Tabletop		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	63%	61%	62%	85%	65%	75%
Dyad 2	84%	46%	65%	94%	69%	82%
Dyad 3	56%	45%	51%	76%	65%	71%
Dyad 4	61%	41%	51%	78%	59%	69%
Dyad 5	56%	78%	67%	77%	93%	85%
Dyad 6	66%	47%	57%	80%	66%	73%
Dyad 8	77%	54%	66%	93%	78%	86%
Dyad 9	66%	45%	56%	85%	66%	76%
Dyad 10	69%	42%	56%	84%	63%	74%
Dyad 13	79%	54%	67%	96%	71%	84%
Dyad 14	43%	42%	43%	62%	66%	64%
Dyad 15	55%	39%	47%	76%	57%	67%
Dyad 16	44%	30%	37%	72%	50%	61%
Dyad 17	62%	36%	49%	84%	59%	72%
Dyad 18	44%	34%	39%	71%	58%	65%
Overall Avg.	62%	46%	54%	81%	66%	73%
Overall S.D.	13%	12%	10%	9%	10%	8%

Table D.1: Average % dyad touch-points for Pop-A-Bubble Single Score Bar (1)

	50 % of Tabletop			66 % of Tabletop		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	58%	41%	50%	80%	74%	77%
Dyad 2	51%	85%	68%	75%	99%	87%
Dyad 3	53%	45%	49%	71%	63%	67%
Dyad 4	59%	51%	55%	79%	68%	74%
Dyad 5	86%	58%	72%	98%	77%	88%
Dyad 6	64%	51%	58%	84%	70%	77%
Dyad 8	69%	52%	61%	85%	71%	78%
Dyad 9	64%	40%	52%	81%	66%	74%
Dyad 10	67%	57%	62%	81%	73%	77%
Dyad 13	58%	71%	65%	74%	90%	82%
Dyad 14	60%	30%	45%	80%	51%	66%
Dyad 15	54%	38%	46%	76%	56%	66%
Dyad 16	47%	48%	37%	64%	66%	65%
Dyad 17	48%	42%	49%	74%	58%	66%
Dyad 18	41%	42%	39%	73%	61%	67%
Overall Avg.	59%	50%	54%	78%	70%	74%
Overall S.D.	11%	14%	10%	8%	12%	8%

Table D.2: Average % Dyad Touch-points for Pop-A-Bubble Single Score Bar (2)

	50 % of Tabletop			66 % of Tabletop		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	57%	44%	51%	80%	66%	73%
Dyad 2	56%	63%	60%	74%	88%	81%
Dyad 3	51%	53%	52%	72%	75%	74%
Dyad 4	46%	65%	56%	66%	83%	75%
Dyad 5	77%	49%	63%	93%	66%	80%
Dyad 6	61%	55%	58%	78%	74%	76%
Dyad 8	71%	72%	72%	89%	90%	90%
Dyad 9	48%	56%	52%	65%	77%	71%
Dyad 10	50%	56%	53%	76%	74%	75%
Dyad 13	65%	64%	65%	81%	87%	84%
Dyad 14	50%	36%	43%	70%	60%	65%
Dyad 15	41%	45%	43%	60%	67%	64%
Dyad 16	53%	29%	41%	76%	55%	66%
Dyad 17	34%	58%	46%	50%	80%	65%
Dyad 18	43%	45%	44%	60%	70%	65%
Overall Avg.	54%	53%	53%	73%	74%	73%
Overall S.D.	11%	12%	9%	11%	10%	8%

Table D.3: Average % Dyad Touch-points for Pop-A-Bubble Dual Score Bar

	50 % of Tabletop			66 % of Tabletop		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	69%	70%	69%	88%	82%	85%
Dyad 2	70%	74%	72%	89%	85%	87%
Dyad 3	62%	58%	60%	85%	84%	85%
Dyad 4	75%	77%	76%	86%	92%	89%
Dyad 5	78%	78%	78%	88%	92%	90%
Dyad 6	74%	70%	72%	87%	80%	83%
Dyad 7	69%	67%	68%	88%	86%	87%
Dyad 8	76%	75%	75%	89%	93%	91%
Dyad 9	58%	45%	52%	80%	74%	77%
Dyad 10	64%	65%	64%	81%	86%	83%
Dyad 11	68%	69%	69%	86%	89%	88%
Dyad 12	84%	78%	81%	93%	90%	92%
Dyad 13	64%	68%	66%	86%	87%	86%
Dyad 14	67%	67%	67%	90%	84%	87%
Dyad 15	73%	67%	70%	86%	88%	87%
Dyad 16	6%	65%	65%	81%	89%	85%
Dyad 17	69%	67%	68%	88%	86%	87%
Dyad 18	69%	73%	71%	86%	90%	88%
Overall Avg.	70%	68%	69%	86%	86%	86%
Overall S.D.	6%	8%	7%	3%	5%	3%

Table D.4: Average % Dyad Touch-points for Collaborative Accuracy Memory

	50 % of Tabletop			**66 % of Tabletop**		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	76%	74%	75%	86%	91%	89%
Dyad 2	78%	80%	79%	91%	87%	89%
Dyad 3	67%	69%	68%	80%	85%	83%
Dyad 4	76%	73%	75%	89%	91%	90%
Dyad 5	77%	73%	75%	90%	95%	92%
Dyad 6	74%	71%	72%	85%	92%	89%
Dyad 7	69%	70%	70%	93%	90%	91%
Dyad 8	74%	73%	74%	90%	93%	91%
Dyad 9	57%	63%	60%	75%	86%	81%
Dyad 10	66%	67%	67%	89%	85%	87%
Dyad 11	73%	71%	72%	92%	91%	91%
Dyad 12	83%	78%	80%	92%	93%	93%
Dyad 13	52%	58%	55%	87%	86%	86%
Dyad 14	76%	75%	76%	89%	94%	91%
Dyad 15	71%	61%	66%	86%	86%	86%
Dyad 16	56%	64%	60%	80%	85%	83%
Dyad 17	69%	68%	68%	88%	91%	89%
Dyad 18	73%	70%	71%	85%	90%	87%
Overall Avg.	70%	70%	70%	87%	89%	88%
Overall S.D.	8%	6%	7%	5%	3%	4%

Table D.5: Average % Dyad Touch-points for Collaborative Speed Memory

	50 % of Tabletop			66 % of Tabletop		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	69%	73%	71%	90%	86%	88%
Dyad 2	61%	65%	63%	87%	88%	87%
Dyad 3	74%	66%	70%	89%	85%	87%
Dyad 4	67%	65%	66%	82%	84%	83%
Dyad 5	62%	70%	66%	84%	95%	89%
Dyad 6	69%	66%	67%	92%	90%	91%
Dyad 7	68%	63%	66%	91%	86%	89%
Dyad 8	66%	67%	66%	84%	84%	84%
Dyad 9	74%	69%	72%	90%	93%	91%
Dyad 10	62%	65%	64%	83%	86%	84%
Dyad 11	62%	64%	63%	82%	89%	85%
Dyad 12	62%	70%	66%	87%	89%	88%
Dyad 13	65%	57%	61%	86%	81%	83%
Dyad 14	68%	63%	66%	87%	90%	88%
Dyad 15	74%	63%	68%	90%	82%	86%
Dyad 16	67%	64%	65%	87%	86%	86%
Dyad 17	69%	70%	69%	90%	92%	91%
Dyad 18	67%	71%	69%	85%	88%	87%
Overall Avg.	67%	66%	66%	87%	87%	87%
Overall S.D.	4%	4%	3%	3%	4%	3%

Table D.6: Average % Dyad Touch-points for Competitive Extravert Memory

	50 % of Tabletop			66 % of Tabletop		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	46%	56%	51%	63%	71%	67%
Dyad 2	46%	80%	63%	65%	88%	77%
Dyad 3	58%	53%	55%	70%	81%	75%
Dyad 4	74%	72%	73%	89%	89%	89%
Dyad 5	53%	79%	66%	66%	94%	80%
Dyad 6	47%	55%	51%	71%	77%	74%
Dyad 7	60%	73%	67%	73%	88%	80%
Dyad 8	63%	77%	70%	83%	92%	88%
Dyad 9	64%	61%	62%	81%	81%	81%
Dyad 10	68%	68%	68%	84%	86%	85%
Dyad 11	47%	78%	63%	73%	86%	79%
Dyad 12	49%	78%	64%	66%	95%	80%
Dyad 13	59%	66%	62%	82%	82%	82%
Dyad 14	58%	71%	65%	79%	87%	83%
Dyad 15	45%	76%	68%	64%	88%	76%
Dyad 16	41%	78%	60%	62%	94%	78%
Dyad 17	54%	76%	65%	77%	87%	82%
Dyad 18	60%	69%	64%	77%	86%	81%
Overall Avg.	55%	70%	63%	74%	86%	80%
Overall S.D.	9%	9%	6%	8%	6%	5%

Table D.7: Average % Dyad Touch-points for Competitive Introvert Memory

	50 % of Tabletop			**66 % of Tabletop**		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	91%	67%	79%	97%	79%	88%
Dyad 2	89%	91%	90%	95%	96%	96%
Dyad 3	89%	73%	81%	95%	85%	90%
Dyad 4	90%	69%	80%	97%	85%	91%
Dyad 5	96%	68%	82%	98%	84%	91%
Dyad 6	90%	74%	82%	96%	88%	92%
Dyad 7	93%	55%	74%	98%	75%	87%
Dyad 8	87%	61%	74%	94%	77%	86%
Dyad 9	91%	45%	68%	98%	65%	82%
Dyad 10	88%	53%	71%	95%	70%	83%
Dyad 11	91%	84%	88%	96%	91%	94%
Dyad 12	88%	72%	80%	95%	82%	89%
Dyad 13	90%	48%	69%	94%	68%	81%
Dyad 14	84%	55%	70%	91%	72%	82%
Dyad 15	92%	65%	79%	97%	80%	89%
Dyad 16	86%	63%	75%	96%	77%	87%
Dyad 17	91%	61%	76%	97%	80%	89%
Dyad 18	86%	8%	47%	93%	91%	92%
Overall Avg.	90%	62%	76%	96%	80%	88%
Overall S.D.	3%	18%	9%	2%	8%	4%

Table D.8: Average % Dyad Touch-points for Físchlár-DT1

	50 % of Tabletop			66 % of Tabletop		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	92%	84%	88%	97%	97%	97%
Dyad 2	92%	95%	93%	97%	96%	97%
Dyad 3	89%	88%	89%	93%	93%	93%
Dyad 4	92%	90%	91%	95%	93%	94%
Dyad 5	91%	95%	93%	95%	97%	96%
Dyad 6	80%	93%	87%	95%	97%	96%
Dyad 7	85%	75%	80%	92%	89%	91%
Dyad 8	90%	82%	86%	95%	91%	93%
Dyad 9	96%	93%	94%	98%	97%	98%
Dyad 10	91%	80%	85%	95%	88%	92%
Dyad 11	86%	94%	90%	92%	97%	95%
Dyad 12	88%	84%	86%	94%	91%	93%
Dyad 13	91%	94%	92%	94%	98%	96%
Dyad 14	89%	88%	89%	95%	96%	95%
Dyad 15	94%	97%	96%	98%	99%	98%
Dyad 16	94%	76%	85%	97%	89%	93%
Dyad 17	89%	80%	84%	94%	89%	91%
Dyad 18	91%	86%	88%	94%	94%	94%
Overall Avg.	90%	87%	89%	95%	94%	94%
Overall S.D.	4%	7%	4%	2%	4%	2%

Table D.9: Average % Dyad Touch-points for Introvert Físchlár-DT2

	50 % of Tabletop			**66 % of Tabletop**		
	User 1	**User 2**	**Average**	**User 1**	**User 2**	**Average**
Dyad 1	84%	91%	88%	92%	98%	95%
Dyad 2	93%	86%	89%	98%	96%	97%
Dyad 3	91%	59%	75%	97%	72%	84%
Dyad 4	96%	61%	78%	98%	73%	86%
Dyad 5	91%	76%	83%	98%	88%	93%
Dyad 6	83%	78%	81%	90%	91%	91%
Dyad 7	89%	62%	75%	96%	77%	87%
Dyad 8	89%	71%	80%	97%	80%	89%
Dyad 9	91%	61%	76%	99%	72%	85%
Dyad 10	90%	63%	77%	98%	76%	87%
Dyad 11	88%	79%	84%	96%	88%	92%
Dyad 12	95%	65%	80%	99%	81%	90%
Dyad 13	90%	66%	78%	94%	80%	87%
Dyad 14	93%	60%	77%	98%	72%	85%
Dyad 15	89%	72%	80%	97%	83%	90%
Dyad 16	69%	38%	53%	88%	58%	73%
Dyad 17	91%	75%	83%	95%	87%	91%
Dyad 18	95%	77%	86%	98%	89%	93%
Overall Avg.	89%	69%	79%	96%	81%	89%
Overall S.D.	6%	12%	8%	3%	10%	5%

Table D.10: Average % Dyad Touch-points for Extravert Físchlár-DT2

50 % of Tabletop						
	Pop-A-Bubble Dual	**Pop-A-Bubble Single(1)**	**Pop-A-Bubble Single(2)**	**Extravert Memory**	**Introvert Memory**	**Avg. of Competitive Systems**
Dyad 1	51%	62%	50%	71%	51%	57%
Dyad 2	60%	65%	68%	63%	63%	64%
Dyad 3	52%	51%	49%	70%	55%	55%
Dyad 4	56%	51%	55%	66%	73%	60%
Dyad 5	63%	67%	72%	66%	66%	67%
Dyad 6	58%	57%	58%	67%	51%	58%
Dyad 7				66%	67%	66%
Dyad 8	72%	66%	61%	66%	70%	67%
Dyad 9	52%	56%	52%	72%	62%	59%
Dyad 10	53%	56%	62%	64%	68%	60%
Dyad 11				63%	63%	63%
Dyad 12				66%	64%	65%
Dyad 13	65%	67%	65%	61%	62%	64%
Dyad 14	43%	43%	45%	66%	65%	52%
Dyad 15	43%	47%	46%	68%	68%	55%
Dyad 16	41%	37%	48%	65%	60%	50%
Dyad 17	46%	49%	42%	69%	65%	54%
Dyad 18	44%	39%	42%	69%	64%	52%
Overall Avg.	53%	54%	54%	66%	63%	59%
Overall S.D.	9%	10%	10%	3%	6%	5%

Table D.11: Average % Dyad Touch-points for Competitive Systems

50 % of Tabletop						
	Accuracy Memory	**Speed Memory**	**Físchlár-DT-1**	**Físchlár-DT-2 Introvert**	**Físchlár-DT-2 Extravert**	**Avg.of Collaborative Systems**
Dyad 1	69%	75%	79%	88%	88%	80%
Dyad 2	72%	79%	90%	93%	89%	85%
Dyad 3	60%	68%	81%	89%	75%	75%
Dyad 4	76%	75%	80%	91%	78%	80%
Dyad 5	78%	75%	82%	93%	83%	82%
Dyad 6	72%	72%	82%	87%	81%	79%
Dyad 7	68%	70%	74%	80%	75%	73%
Dyad 8	75%	74%	74%	86%	80%	78%
Dyad 9	52%	60%	68%	94%	76%	70%
Dyad 10	64%	67%	71%	85%	77%	73%
Dyad 11	69%	72%	88%	90%	84%	80%
Dyad 12	81%	80%	80%	86%	80%	81%
Dyad 13	66%	55%	69%	92%	78%	72%
Dyad 14	67%	76%	70%	89%	77%	75%
Dyad 15	70%	66%	79%	96%	80%	78%
Dyad 16	65%	60%	75%	85%	53%	68%
Dyad 17	68%	68%	76%	84%	83%	76%
Dyad 18	71%	71%	47%	88%	86%	73%
Overall Avg.	69%	70%	76%	89%	79%	76%
Overall S.D.	7%	7%	9%	4%	8%	5%

Table D.12: Average % Dyad Touch-points for Collaborative Systems

66 % of Tabletop						
	Pop-A-Bubble Dual	**Pop-A-Bubble Single(1)**	**Pop-A-Bubble Single(2)**	**Extravert Memory**	**Introvert Memory**	**Avg. of Competitive Systems**
Dyad 1	73%	75%	77%	88%	67%	76%
Dyad 2	81%	82%	87%	87%	77%	83%
Dyad 3	74%	71%	67%	87%	75%	75%
Dyad 4	75%	69%	74%	83%	89%	78%
Dyad 5	80%	85%	88%	89%	80%	84%
Dyad 6	76%	73%	77%	91%	74%	78%
Dyad 7				89%	80%	84%
Dyad 8	90%	86%	78%	84%	88%	85%
Dyad 9	71%	76%	74%	91%	81%	79%
Dyad 10	75%	74%	77%	84%	85%	79%
Dyad 11				85%	79%	82%
Dyad 12				88%	80%	84%
Dyad 13	84%	84%	82%	83%	82%	83%
Dyad 14	65%	64%	66%	88%	83%	73%
Dyad 15	64%	67%	66%	86%	76%	72%
Dyad 16	66%	61%	65%	86%	78%	71%
Dyad 17	65%	72%	66%	91%	82%	75%
Dyad 18	65%	65%	67%	87%	81%	73%
Overall Avg.	73%	73%	74%	87%	80%	78%
Overall S.D.	8%	8%	8%	3%	5%	5%

Table D.13: Average % Dyad Touch-points for Competitive Systems

66 % of Tabletop						
	Accuracy Memory	**Speed Memory**	**Físchlár-DT-1**	**Físchlár-DT-2 Introvert**	**Físchlár-DT-2 Extravert**	**Avg. of Collaborative Systems**
Dyad 1	85%	89%	88%	97%	95%	91%
Dyad 2	87%	89%	96%	97%	97%	93%
Dyad 3	85%	83%	90%	93%	84%	87%
Dyad 4	89%	90%	91%	94%	86%	90%
Dyad 5	90%	92%	91%	96%	93%	92%
Dyad 6	83%	89%	92%	96%	91%	90%
Dyad 7	87%	91%	87%	91%	87%	89%
Dyad 8	91%	91%	86%	93%	89%	90%
Dyad 9	77%	81%	82%	98%	85%	84%
Dyad 10	83%	87%	83%	92%	87%	86%
Dyad 11	88%	91%	94%	95%	92%	92%
Dyad 12	92%	93%	89%	93%	90%	91%
Dyad 13	86%	86%	81%	96%	87%	87%
Dyad 14	87%	91%	82%	95%	85%	88%
Dyad 15	87%	86%	89%	98%	90%	90%
Dyad 16	85%	83%	87%	93%	73%	84%
Dyad 17	87%	89%	89%	91%	91%	89%
Dyad 18	88%	87%	92%	94%	93%	91%
Overall Avg.	86%	88%	88%	94%	89%	89%
Overall S.D.	3%	4%	4%	2%	5%	3%

Table D.14: Average % Dyad Touch-points for Collaborative Systems

Bibliography

ACE (2008). Url: http://sourceforge.net/projects/ace.

Adcock, J., Girgensohn, A., Cooper, M., Liu, T., Wilcox, L., and Rieffel, E. (2004). Fxpal experiments for trecvid 2004. In *TRECVID 2004: - Text REtrieval Conference TRECVID Workshop*.

Agnihotri, L., Kender, J., Dimitrova, N., and Zimmerman, J. (2005). User study for generating personalized summary profiles. In *IEEE International Conference on Multimedia and Expo, 2005*, pages 1094–1097.

Alertus®(2008). Alertus® homepage. In *URL: http://www.alertus.com/index.php*.

Allport, G. (1937). *Personality; a psychological interpretation.* H. Holt and Co., New York.

Allport, G. and Odbert, H. (1936). *Trait-names, a psycho-lexical study.* Psychological Review Co., Princeton, N.J., Albany, N.Y.

American-Heritage-Dictionary (2008). Url: http://dictionary.reference.com/browse/guideline. In *The American Heritage® Dictionary*.

Anderson, R., Bikson, T., Law, S., and Mitchell, B. (2001). Universal access to email: feasibility and societal implications. pages 243–262.

Baecker, R., Moore, G., and Zijdemans, A. (2003). Reinventing the lecture:

Webcasting made interactive. In *Proceedings of HCI International*, pages 896–900. Lawrence Erlbaum Associates.

Baeza-Yates, R. and Ribeiro-Neto, B. (2008). *Modern Information Retrieval*. Pearson Education Canada.

Balthazard, P., Potter, R., and Warren, J. (2004). Expertise, extraversion and group interaction styles as performance indicators in virtual teams: how do perceptions of it's performance get formed? *ACM SIGMIS Database*, 35(1):41–64.

Bebo (2008). Url: http://www.bebo.com/.

Berger, A. (2002). *Video Games: A Popular Culture Phenomenon*. Transaction Publishers.

Bier, E., Freeman, S., and Pier, K. (1992). Mmm: The multi-device multi-user multi-editor. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 645–646, Monterey, California, United States. ACM.

Billinghurst, M., Grasset, R., Green, R., and Haller, M. (2005). Inventing the future down under: the human interface technology laboratory new zealand [hit lab nz]. *Multimedia Systems*, 39(2):18–23.

Bly, S. A., Harrison, S., and Irwin, S. (1993). Media spaces: bringing people together in a video, audio, and computing environment. *Commun. ACM*, 36(1):28–46.

Borella, M. (2000). Source models of network game traffic. *Computer Communications*, 23(4):403–410.

Borgatta, E. (1964). The structure of personality characteristics. *Behavioral Science*, 12(1):8–17.

Borlund, P. (2003). The iir evaluation model: a framework for evaluation of interactive information retrieval systems. *Journal of Information Research*, 8(3):paper no. 152.

Bottger, P. (1984). Expertise and air time as bases of actual and perceived influence in problem-solving groups. *Journal of Applied Psychology*, 69(2):214–221.

Brewer, T. and Bowcock, J. (2008). Apple introduces the new iphone 3g. In *URL: http://www.apple.com/pr/library/2008/06/09iphone.html*.

Briggs Myers, I. and Hauley McCauley, M. (1985). *MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type, 5th edition*. Consulting Psychologists Press.

Brinkman, W.-P. and Fine, N. (2005). Towards customized emotional design: an explorative study of user personality and user interface skin preferences. In *Proceedings of the 2005 annual conference on European association of cognitive ergonomics*, volume 132, pages 107–114. University of Athens.

Browne, P., Czirjek, C., Gaughan, G., Gurrin, C., Jones, G., Lee, H., Marlow, S., McDonald, K., Murphy, N., O'Connor, N., O'Hare, N., Smeaton, A., and Ye, J. (2003). Dublin City University video track experiments for TREC 2003. In *TRECVID 2003: - Text REtrieval Conference TRECVID Workshop*.

Browne, P., Czirjek, C., Gaughan, G., Gurrin, C., Jones, G., Lee, H., Marlow, S., McDonald, K., Murphy, N., O'Connor, N., O'Hare, N., Smeaton, A., and Ye, J. (2004). Dublin City University video track experiments for TREC 2004. In *TRECVID 2004: - Text REtrieval Conference TRECVID Workshop*.

Byrne, D. and Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of personality and social psychology*, 4:240–243.

Byström, K. and Järvelin, K. (1995). Task complexity affects information seeking and use. *Inf. Process. Manage.*, 31(2):191–213.

Card, S., Moran, T., and Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Lawrence Erlbaum Associates, Publishers, New Jersey.

Castillo, J. and Hartson, H. (2007). Remote usability testing methods a la carte. Technical report, Computer Science, Virginia Tech, USA.

Castronova, E. (2001). Virtual worlds: A first-hand account of market and society on the cyberian frontier. *CESifo Working Paper Series*, (No. 618).

Cattell, R. (1943). The description of personality: Basic traits resolved into clusters. *Abnormal and social psychology*, 38(426):506.

Cattell, R. (1945a). The description of personality: Principles and findings in a factor analysis. *The American journal of psychology*, 58(1):69–90.

Cattell, R. (1945b). The principal trait clusters for describing personality. *Psychological bulletin*, 42(3):129–161.

Cattell, R., Eber, H., and Tatsuoka, M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16 PF).* Institute for Personality and Ability Testing.

Chaplin, W., John, O., and Goldberg, L. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Personality and Social Psychology*, 54(4):541–557.

Cleverdon, C. and Keen, E. (1966). Aslib cranfield research project - factors determining the performance of indexing systems. Technical report, College of Aeronautics, Cranfield.

Cleverdon, C., Mills, J., and Keen, E. (1966). Aslib cranfield research project - factors determining the performance of indexing systems. Technical report, College of Aeronautics, Cranfield.

Cohen, J., Cohen, P., West, S., and Aiken, L. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Lawrence Erlbaum Associates.

Coleman, D. and Willeges, R. (1985). Collecting detailed user evaluations of software interface. In *In Proceedings of Human Factors Society, Twenty-ninth annual meeting*, pages 240–244.

Cooke, R. (1994). *The Ethical Decision Challenge*. Human Synergistics/Center for Applied Research.

Cooke, R. and Kernaghan, J. (1987). Estimating the difference between group versus individual performance on problem-solving tasks. *Group Organization Management*, 12(3):319–342.

Cooley, C. (1918). *Social Process*. Charles Scribner's Sons.

Costa, P. and McCrae, R. (1993). Bullish on personality psychology. *The Psychologist*, 6:302–303.

Costa, P. and McCrea, R. (1992a). *NEO PI-R Professional Manual*. Psychological Assessment Resources.

Costa, P. and McCrea, R. (1992b). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources.

Costa, P. T. and McCrae, R. R. (1976). Age differences in personality structure: A cluster analytic approach. *Journal of Gerontology*, 31(5):564–570.

Da Cunha, A. and Greathead, D. (2007). Does personality matter? an analysis of code-review ability. *Communications of the ACM*, 50(5):109–112.

DARPA (2008). Url: http://www.darpa.mil/.

De Raad, B., Perugini, M., Hrebckov, M., and Szarota, P. (1998). Lingua franca of personality. *Journal of Cross-Cultural Psychology*, 29(1):212–232.

Dictionary-Of-Computing (2008). Url: http://dictionary.reference.com/browse/evaluation. In *The Free On-line Dictionary of Computing*.

Dietz, P. and Leigh, D. (2001). Diamondtouch: a multi-user touch technology. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226, Orlando, Florida. ACM.

Digman, J. and Takemoto-Chock, N. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate behavioral research*, 16(2):149–170.

Doan, D., Paton, N., and Kilgour, A. (1995). Design and user testing of a multi-paradigm query interface to an object-oriented database. *SIGMOD Rec.*, 24(3):12–17.

DocSynch (2008). Url: http://docsynch.sourceforge.net/.

Ehrlich, S. (1987). Strategies for encouraging successful adoption of office communication systems. *ACM Trans. Inf. Syst.*, 5(4):340–357.

Ellis, C. and Gibbs, S. (1989). Concurrency control in groupware systems. *SIGMOD Rec.*, 18(2):399–407.

Ellis, C., Gibbs, S., and Rein, G. (1991). Groupware: some issues and experiences. *Communications of the ACM*, 34(1):39–58.

Ellis, D. (1989). A behavioral approach to information retrieval system design. *Journal of Documentation*, 45(3):171–212.

Erikson, E. (1969). *Gandhi's Truth: On the Origins of Militant Nonviolence*. W.W. Norton & Company.

Eysenck, H. (1947). *Dimensions of Personality*. Praegar.

Eysenck, H. and Eysenck, S. (1964). *Manual of Eysenck Personality Inventory*. University Press.

Facebook (2008). Url: http://www.facebook.com/.

Fhios (2008). Fhios homepage. In *URL: http://www.fhios.com/*.

Fine, N. and Brinkman, W. (2006). Towards extracting personality trait data from interaction behaviour. In *Position Paper in Human-Computer Interaction Engage Workshop*. Springer.

Fiske, D. (1949). Consistency of the factorial structures of personality ratings from different sources. *Abnormal and Social Psychology*, 44.

Fitts, P. (1994). The information capacity of the human motor system in controlling the amplitude of movement. *Experimental Psychology*, 47(1):381–391.

Foley, C., Gurrin, C., Jones, G., Lee, H., Mc Givney, S., OConnor, N., Sav, S., Smeaton, A., and Wilkins, P. (2005). Trecvid 2005 experiments at dublin City University. In *TRECVID 2005: - Text REtrieval Conference TRECVID Workshop*.

Forlines, C., Esenther, A., Shen, C., Wigdor, D., and Ryall, K. (2006). Multi-user, multi-display interaction with a single-user, single-display geospatial application. In *UIST '06: Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 273–276, Montreux, Switzerland. ACM.

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. SAGE.

Freidman, H. and Schustack, M. (2006). *Personality: Classic Theories and Modern Research, 3rd Edition*. Pearson.

Freud, S. and Riviere, J. (1927). *The Ego and the Id*. L. & Virginia Woolf at the Hogarth press, and the Institute of psycho-analysis.

Freye, J., Varga, E., Byrne, D., Smeaton, A., Smyth, B., and Jones, G. (2007). Realising context-sensitive mobile messaging. In *MONET'07 - 2nd International Workshop on MObile and NEtworking Technologies for Social Applications*, volume 4805, pages 407–416. Springer-Verlag.

Gardner, W. and Martinko, M. (1996). Using the myers-briggs type indicator to study managers: A literature review and research agenda. *Journal of Management*, 22(1):45–83.

Goldberg, L. (1981). Language and individual differences. the search for universals in personality lexicons. *Personality and Social Psychology Review*, 2:141–165.

GOOGLE (2008a). Url: http://www.google.com/calendar.

GOOGLE (2008b). Url: http://www.groups.google.com/.

Gorla, N. and Lam, Y. (2004). Who should work with whom?: building effective software project teams. *Communications of the ACM*, 47(6):79–82.

Groove (2008). Url: http://www.microsoft.com/presspass/features/2005/mar05/03-10grooveqa.mspx.

Grudin, J. (1994). Eight challenges for developers. *Communications of the ACM*, 37(1):92–105.

GSM (1985). Services and facilities to be provided in the gsm system, rev2. In *GSM Doc 28/85*.

GTalk (2008). Url: http://www.google.com/talk/.

Gurrin, C., Browne, P., Smeaton, A., Lee, H., Mc Donald, K., and P., M. (2004). Físchlár-nursing, using digital video libraries to teach processes to nursing students. In *WBE 2004 - IASTED International Conference on Web-Based Education*, pages 111–116.

Gutwin, C. and Greenberg, S. (1998). Design for individuals, design for groups: Tradeoffs between power and workspace awareness. In *CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 207–216, Seattle, Washington, United States. ACM.

Gutwin, C., Roseman, M., and Greenberg, S. (1996). A usability study of awareness widgets in a shared workspace groupware system. In *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 258–267, Boston, Massachusetts, United States. ACM.

Hackman, J. and Morris, C. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in Experimental Social Psychology*.

Han, J. (2005). Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118, Seattle, WA, USA. ACM.

Hans, J. (2008). Multi-Touch Photo. In *URL: http://cs.nyu.edu/ jhan/ftirtouch/index.html*.

Hansen, W. (1971). User engineering principles for interactive systems. In *Fall Joint Computer Conference*, pages 523–532. AFIPS Press.

Hartson, H. and Castillo, J. (1998). Remote evaluation for post-deployment usability improvement. In *AVI '98: Proceedings of the working conference on Advanced visual interfaces*, pages 22–29, New York, NY, USA. ACM.

Heesch, D., Howarth, P., Megalhaes, J., May, A., Pickering, M., Yavlinsky, A., and Ruger, S. (2004). Video retrieval using search and browsing. In *TRECVID 2004: - Text REtrieval Conference TRECVID Workshop*.

Henry-Labordere, A. and Jonack, V. (2004). *SMS and MMS Interworking in Mobile Networks*. Artech House.

Herbsleb, J., Atkins, D., Boyer, D., Handel, M., and Finholt, T. (2002). Introducing instant messaging and chat in the workplace. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 171–178, New York, NY, USA. ACM.

Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, Strong, and Verplank (2008). Url: http://www.sigchi.org/cdg/cdg2.html#2_1.

Hutterer, P. and Thomas, B. (2007). Groupware support in the windowing system. In *Proceedings of the eight Australasian conference on User interface*, volume 64, pages 39–46, Ballarat, Victoria, Australia.

Iacono, S. and Weisband, S. (1997). Developing trust in virtual teams. In *HICSS '97: Proceedings of the 30th Hawaii International Conference on System Sciences*, pages 412–420, Washington, DC, USA. IEEE Computer Society.

Ishii, H., Kobayashi, M., and Grudin, J. (1992). Integration of inter-personal space and shared workspace: Clearboard design and experiments. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 33–42. ACM.

Jarvenpaa, S., Knoll, K., and Leidner, D. (1988). Is anybody out there? antecedents of trust in global virtual teams. *Management Information Systems*, 14(4):29–64.

Johnson, A. (2008). Short Form for the IPIP-Neo (International Personality Item Pool Representation of the NEO PI-R)). In *URL: http://www.personal.psu.edu/faculty/j/5/j5j/IPIP/ipipneo120.htm.*

Jung, C. (1971). *Psychological Types.* Princeton University Press.

Jung, C. and Adler, G. c. (1969). *The Archetypes and the Collective Unconscious.* Princeton University Press.

Jung, C. and Baynes, H. (1926). *Psychological Types: Or, The Psychology of Individuation.* Harcourt, Brace & company, inc.

Kakehi, Y., Hosomi, T., Iida, M., Naemura, T., and Matsushita, M. (2006). Transparent tabletop interface for multiple users on lumisight table. In *TABLETOP '06: Proceedings of the First IEEE International Workshop* *on Horizontal Interactive Human-Computer Systems*, pages 143–150, Adelaide, Australia. IEEE Computer Society.

Karsvall, A. (2002). Personality preferences in graphical interface design. In *NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction*, pages 217–218, Aarhus, Denmark. ACM.

Keirsey, D. (2008). Url: http://www.keirsey.com/.

Kelly, G. (1963). *Theory Of Personality: The Psychology of Personal Constructs*. W.W. Norton.

Kender, J., Lin, C.-Y., Naphade, M., Natsev, A., Smith, J., sić J., T., Wu, G., Yan, R., Zhang, D., Argillander, J., Franz, M., Iyengar, G., Amir, A., Berg, M., Chang, S.-F., Hsu, W., and Kennedy, L. (2004). Ibm research trecvid 2004 video retrieval system. In *TRECVID 2004: - Text REtrieval Conference TRECVID Workshop*.

Kennedy, R. and Kennedy, D. (2004). Using the myers-briggs type indicator in career counseling. *Journal of Employment Counseling*, 41(1):38–44.

Kraaij, W. and Post, W. (2006). Task-based evaluation of exploratory search systems. In *In Proceedings of SIGIR 2006 workshop, Evaluating Exploratory Search Systems*. ACM Press.

Kruger, R., Carpendale, S., Scott, S., and Greenberg, S. (2003). How people use orientation on tables: comprehension, coordination and communication. In *GROUP '03: Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, pages 369–378, Sanibel Island, Florida, USA. ACM.

Kuhlthau, C. (1992). Seeking meaning: a process approach to library and information services. Ablex Publishing.

Kwasnik, B. (1992). A descriptive study of the functional components of browsing. In *Proceedings of the IFIP TC2/WG2.7 Working Conference on En-*

gineering for Human-Computer Interaction, pages 191–203, Amsterdam, The Netherlands, The Netherlands. North-Holland Publishing Co.

Lafferty, J. and Eady, P. (1974). *The desert survival problem.* Plymouth, MI: Experimental Learning Methods.

Lancaster, F. (1969). Medlars: report on the evaluation of its operating efficiency. *American Documentation*, 10(2):119–142.

Lee, H., Smeaton, A., O'Connor, N., and Smyth, B. (2006). User evaluation of fschlr-news: An automatic broadcast news delivery system. *TOIS - ACM Transactions on Information Systems*, 24(2):145–189.

Lee, H., Smeaton, A., O'Toole, C., Murphy, N., Marlow, S., and O'Connor, N. (2000). The Físchlár digital video recording, analysis, and browsing system. In *RIAO 2000: Content-based Multimedia Information Access, Paris, France*, pages 26–28.

Linden (2008). Second life. In *URL: http://secondlife.com/.*

Lotus (2008). Url: http://www-128.ibm.com/developerworks/lotus/library/ls-ndhistory/.

Ludford, P., Cosley, D., Frankowski, D., and Terveen, L. (2004). Think different: increasing online community participation using uniqueness and group dissimilarity. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 631–638, Vienna, Austria. ACM.

Marchionini, G. (1995). *Information seeking in electronic environments.* Cambridge University Press.

Marchionini, G. and Shneiderman, B. (1988). Finding facts vs. browsing knowledge in hypertext systems. *Computer*, 21(1):70–80.

McDonald, S. and Edwards, H. (2007). Who should test whom. *Communications of the ACM*, 50(1):67–71.

Mead, G., Morris, C., Brewster, J., Dunham, A., and Miller, D. (1938). *The Philosophy of the Act*. The University of Chicago press.

Meet-O-Matic (2008). Url: http://www.meetomatic.com/calendar.php.

Merriam-Webstar (2008). http://www.merriam-webster.com/dictionary/study. In *The Merriam-Webstar Dictionary*.

Merrill-Reed (2008). Url: http://www.tracomcorp.com/article/ca413073.html.

Microsoft (2007). Url: http://www.microsoft.com/presspass/press/2007/may07/05-29mssurfacepr.mspx.

Microsoft (2008a). Url: http://messengersays.spaces.live.com/blog/cns!5b410f7fd930829e!29966.entry.

Microsoft (2008b). Xbox 360 ushers in the future of games and entertainment. In *URL: http://www.microsoft.com/presspass/press/2005/may05/05-12360MTVUnveilPR.mspx*.

Müller-Tomfeld, C. and Reischl, W. (1998). Communication chairs: examples of mobile roomware components. In *CHI '98: CHI 98 conference summary on Human factors in computing systems*, pages 267–268, New York, NY, USA. ACM.

Nardi, B., Whittaker, S., and Bradner, E. (2000). Interaction and outeraction: instant messaging in action. In *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 79–88, New York, NY, USA. ACM.

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts.

New-Launches (2007). Url: http://www.newlaunches.com/archives/surface_based_computer_for_mac_fans.php.

Nielsen, J. (1993). *Usability Engineering*. Morgan Kauffman.

Nielsen, J. and Mack, R. E. (1994). *Usability Inspection Methods*. John Wiley & Sons, New York, NY.

Nintendo (2008a). Nintendo wii press release. In *URL: http://www.wiiplus.net/fullNews.php?newsId=93*.

Nintendo, D. (2008b). Url: http://www.nintendo.com/ds.

NIST (2008). Url: http://www.nist.gov/.

Norman, D. (1998). *The Design of Everyday Things*. MIT Press, London.

Norman, W. (1963). Toward an adequate taxonomy of personality attributes: replicated factors structure in peer nomination personality ratings. *Abnormal and social psychology*, 66(6):574–90.

Norman, W. (1967). 2,800 personality trait descriptors: Normative operating characteristics for a university population.

Northrop (2008). An immersive visualization and collaboration environment. In *http://www.ms.northropgrumman.com/images/TouchTable_FS.pdf*.

O'Brien, J. (2008). Url: http://www.justthedesignguy.com/.

OED-Online (2008). Url: http://dictionary.oed.com/. In *Oxford English Dictionary Online*.

Pervin, L. and John, O. (1999). *Handbook of Personality: Theory and Research, 2nd Edition*. The Guilford Press.

Pickford, R. (1972). *Psychology and Visual Aesthetics*. Hutchinson Educational LTD, London.

Reeves, B. and Nass, C. (1996). *The Media Equation. Chapter on Media and Personality, page 75*. CSLI Publications.

Rekimoto, J. (2002). Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120, Minneapolis, Minnesota, USA. ACM.

Ringel Morris, M., Morris, D., and Winograd, T. (2004a). Individual audio channels with single display groupware: effects on communication and task strategy. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 242–251, Chicago, Illinois, USA. ACM.

Ringel Morris, M., Paepcke, A., Winograd, T., and Stamberger, J. (2006). Teamtag: exploring centralized versus replicated controls for co-located tabletop groupware. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1273–1282, Montréal, Québec, Canada. ACM.

Ringel Morris, M., Ryall, K., Shen, C., Forlines, C., and Vernier, F. (2004b). Beyond "social protocols": multi-user coordination policies for co-located groupware. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 262–265, Chicago, Illinois, USA. ACM.

Robertson, S. and Hancock-Beaulieu, M. (1992). On the evaluation of ir systems. *Information Processing & Management*, 28(4):457–466.

Rogers, C. (1959). A theory of therapy, personality, and interpersonal relationships, as developed in the client-centered framework. *In S. Koch (Ed.), A study of a science: Study 1. Conceptual and systematic: Vol. 3 Formulations of the person and the social context*, pages 184–256.

Rutherfoord, R. (2006). Using personality inventories to form teams for class projects: a case study. In *SIGITE '06: Proceedings of the 7th conference on Information technology education*, pages 9–14, Minneapolis, Minnesota, USA. ACM.

Ryall, K., Forlines, C., Shen, C., and Morris, M. (2004). Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. In *CSCW '04: Proceedings of the 2004 ACM conference on*

Computer supported cooperative work, pages 284–293, Chicago, Illinois, USA. ACM.

Saati, B., Salem, M., and Brinkman, W.-P. (2005). Towards customized user interface skins: investigating user personality and skin colour. In *HCI2005*, volume 2, pages 86–93, Edinburgh, Scotland.

Saracevic, T. and Kantor, P. (1988a). A study of information seeking and retrieving. ii. users, questions and effectiveness. *Journal of American Society for Information Science*, 39(3):177–196.

Saracevic, T. and Kantor, P. (1988b). A study of information seeking and retrieving. iii. searchers, searches, and overlap. *Journal of American Society for Information Science*, 39(3):197–216.

Sayles, K. and Novick, D. (2004). Assessing effectiveness of personality style in documentation. In *SIGDOC '04: Proceedings of the 22nd annual international conference on Design of communication*, pages 75–82, Memphis, Tennessee, USA. ACM.

Scott, S., Carpendale, M., and Inkpen, K. (2004). Territoriality in collaborative tabletop workspaces. In *ACM conference on Computer Supported Cooperative Work (CSCW)*, pages 294–303, Chicago, Illinois, USA. ACM.

Scott, S., Grant, K., and Mandryk, R. (2003). System guidelines for co-located, collaborative work on a tabletop display. In *ECSCW'03: Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work*, pages 159–178, Norwell, MA, USA. Kluwer Academic Publishers.

Shell, M. M. and Duncan, S. (2000). Url: http://clearinghouse.missouriwestern.edu/manuscripts/177.asp.

Shen, C., Vernier, F., Forlines, C., and Ringel, M. (2004). Diamondspin: an extensible toolkit for around-the-table interaction. In *CHI '04: Proceedings* *of the SIGCHI conference on Human factors in computing systems*, pages 167–174, Vienna, Austria. ACM.

Shim, J. and Paul, B. (2007). Effects of personality types on the use of television genre. *Journal of Broadcasting & Electronic Media*, 51(2):287–304.

Shneiderman, B. and Plaisant, C. (2005). *Designing the User Interface, International Edition*. Addison-Wesley.

Smeaton, A. (2007). Techniques used and open challenges to the analysis, indexing and retrieval of digital video. *Information Systems*, 32(4):545–559.

Smeaton, A., Foley, C., Byrne, D., and Jones, G. (2008). ibingo mobile collaborative search. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 547–548, New York, NY, USA. ACM.

Smeaton, A., Foley, C., Gurrin, C., Lee, H., and Mc Givney, S. (2006a). Collaborative searching for video using the fschlr system and a diamondtouch table. In *TABLETOP '06: Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, pages 151–159. IEEE Computer Society.

Smeaton, A., Kraaij, W., and Over, P. (2005). TRECVID 2005 - an overview. In *TRECVID 2005 - Text REtrieval Conference TRECVID Workshop*.

Smeaton, A., Lee, H., Foley, C., Gurrin, C., and Mc Givney, S. (2006b). Físchlár-diamondtouch: Collaborative video searching on a table. In *SPIE Electronic Imaging - Multimedia Content Analysis, Management, and Retrieval, SPIE Vol. 6073*.

Smeaton, A., Lee, H., Foley, C., and Mc Givney, S. (2007). Collaborative video searching on a tabletop. *Multimedia Systems*, 12(4-5):375–391.

Smeaton, A., Lee, H., and McDonald, K. (2004). Experiences of creating four video library collections with the físchlár system. *Int. J. Digit. Libr.*, 4(1):42–44.

Smilemania (2008). Url: http://www.smilemania.com/wst_page14.html. In *Break Out the Crayons it's a Crises*.

Sony (2008). Url: http://www.scei.co.jp/corporate/data/bizdatajpn_e.html.

Sparck Jones, K. and Willett, P., editors (1997). *Readings in information retrieval*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Stewart, J., Raybourn, E., Bederson, B., and Druin, A. (1998). When two hands are better than one: Enhancing collaboration using single display groupware. In *CHI '98: CHI 98 conference summary on Human factors in computing systems*, pages 287–288, Los Angeles, California, United States. ACM.

Streitz, N., GeiBler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W., Rexroth, P., Seitz, P., and Steinmetz, R. (1999). i-land: an interactive landscape for creativity and innovation. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 120–127, Pittsburgh, Pennsylvania, United States. ACM.

SynchroEdit (2008). Url: http://www.synchroedit.com/.

Tandler, P. (2004). The beach application model and software framework for synchronous collaboration in ubiquitous computing environments. *J. Syst. Softw.*, 69(3):267–296.

Tandler, P., Prante, T., Müller-Tomfelde, C., Streitz, N., and Steinmetz, R. (2001). Connectables: dynamic coupling of displays for the flexible creation of shared workspaces. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 11–20, Orlando, Florida. ACM.

TatukGIS (2008). Url: http://www.tatukgis.com.

Teuscher, C. and Hofstadter, D. (2004). *Alan Turing: Life and Legacy of a Great Thinker. P.381*. Springer.

Toney, A. and Thomas, B. (2006). Applying reach in direct manipulation user interfaces. In *OZCHI '06: Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments*, pages 393–396, Sydney, Australia. ACM.

TRECVid-Guidelines (2005). Url: http://www-nlpir.nist.gov/projects/tv2005/tv2005.html.

Tse, E., Histon, J., Scott, S., and Greenberg, S. (2004). Avoiding interference: how people use spatial separation and partitioning in sdg workspaces. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 252–261, Chicago, Illinois, USA. ACM.

Tse, E., Shen, C., Greenberg, S., and Forlines, C. (2006). Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. In *AVI '06: Proceedings of the working conference on Advanced visual interfaces*, pages 336–343, Venezia, Italy. ACM.

Tuddenham, P. (2007). Distributed tabletops: territoriality and orientation in distributed collaboration. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, pages 2237–2242, San Jose, CA, USA. ACM.

Tupes, E. and Christal, R. (1961). *Recurrent Personality Factors Based on Trait Ratings*. Lackland Air Force Base.

Urdan, T. (2005). *Statistics in Plain English*. Routledge.

Usenet.com (2008). Url: http://www.usenet.com/.

Van Rijsbergen, C. (1979). *Information Retrieval*. Butterworths.

Vaughan Nichols, S. (2007). New interfaces at the touch of a fingertip. *Computer*, 40(8):12–15.

View, M. S.-. (2008). Url: http://www.microsoft.com/surface.

Villa, R., Gildea, N., and Jose, J. (2008). A study of awareness in multimedia search. In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pages 221–230, Pittsburgh, PA, USA. ACM.

Voorhees, E. (2004). Overview of TREC 2004. In *In Proceedings of the 13th Text Retrieval Conference (TREC 2004)*, pages 26–28.

Voorhees, E. and Harman, D. (2003). Erratum to letters to the editor. *Information Processing & Management*, 39(3):503–506.

Walczak, M., McAllister, M., and Segen, J. (2004). Dialog table. In *DIS '04: Proceedings of the 5th conference on Designing interactive systems*, pages 311–311, Cambridge, MA, USA. ACM.

Watson, J. (1929). *Psychology from the Standpoint of a Behaviorist*. J.B. Lippincott company, Original from the University of Michigan.

Westerveld, T., de Vries, A., Ianeva, T., Boldareva, L., and Hiemstra, D. (2003). Combining information sources for video retrieval. In *TRECVID 2003: - Text REtrieval Conference TRECVID Workshop*.

Wilson, T. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3):249–270.

Zillmann, D. and Weaver, J. (1997). Psychoticism in the effect of prolonged exposure to gratuitous media violence on the acceptance of violence as a preferred means of conflict resolution. *Personality and Individual Differences*, 22(5):613–627.