QUALITY IN USE: ADDRESSING AND VALIDATING AFFECTIVE REQUIREMENTS

(Brian) Todd Bentley

For fulfilment of a PhD

Submitted: Dec 2006

Swinburne University of Technology, Melbourne, Australia
Abstract

From a user's perspective, product quality has typically been equated to usability, where the primary focus is performance (i.e. efficiency and effectiveness) and user satisfaction. This view on quality has been adequate to validate the 'fitness for purpose' for many products. However, many other products, such as computer games or at-home-technologies, are not based solely on performance-based qualities. By defining and validating the quality of these non-performance-based technologies using the same measures as performance-based technologies fails to address a key quality – the ability of the product to create fun, enjoyment, or other user affective experiences.

This thesis investigates system quality issues in non-performance-based technologies, primarily addressing affect. It explores what affect is in relation to system quality, how can affect impact quality in use, what product characteristics can aid in the creation of positive user affect, and how to validate whether a product successfully created the desired user affect.

To begin, current knowledge about the relationship between systems development, quality in use, and affect is summarised. Concepts used synonymously with affect were studied (e.g. satisfaction, user experience, emotion). It was found that these are related to affect, but are not affect themselves. Following this, a broad survey asking users to identify product characteristics that create positive affect during use is reported on. Results showed that the system characteristics of learnability, feedback, curiosity, and user cooperation can each increase and enhance positive user affect experienced.

If it is an aim for the system to create an affective experience in a user then this should be considered a requirement of the system; and as a requirement, it is necessary to validate that it achieves this aim. To this end, a laboratory study exploring psychophysiology and cued-recall debrief as methods to evaluate user affect was conducted. Results revealed that both methods show promise for evaluating user affect.
recall is capable of identifying specific user affects, but is limited because it relies on the user to report these affects. Physiological measures can objectively identify when a person is experiencing affect, but is limited because it is not possible to determine what affect is being experienced. Combined, these methods represent an optimal and viable evaluation method for user affect.
Acknowledgements

My sincere thanks to my advisors and friends, Lorraine Johnston and Karola von Baggo, who guided me through numerous obstacles in order to complete this thesis. I believe that without their motivation, encouragement, support, and advice this thesis would never have been completed. I also apologise for all the stress that I am sure I caused (and perhaps still am…).

My appreciation also extends to all who contributed, discussed, dissected, and generally made me feel that the study of affect and quality in use is worthwhile.

Finally - my family. I would like to thank my parents for supporting me and offering their encouragement throughout. They provided encouragement throughout, and were supportive despite meaning a lot less time to see them. And my deepest gratitude goes to my wife Jemima for putting up with the countless hours and stressful nights / weekends to allow me to complete this work. I guess I no longer have an excuse to get out of housework…

Thank-you all,

Todd Bentley
Declaration

This thesis contains no materials which have been previously accepted for the award of any other degree or diploma, except where due reference is made within the text of the thesis. To the best of my knowledge this thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis.

Signed: ______________________
  Todd Bentley

Date: ______________________
# Table of Contents

LIST OF FIGURES .................................................................................................................. X

LIST OF TABLES ...................................................................................................................... XII

CHAPTER 1  INTRODUCTION ................................................................................................. 1

  1.1. THE RESURGENCE OF INTEREST IN AFFECT AND PRODUCT DESIGN ................ 2
  1.2. IMPACTS OF AFFECT ON QUALITY IN USE ............................................................ 4
  1.3. RESEARCH OBJECTIVES ............................................................................................. 5
  1.4. THESIS ORGANISATION .............................................................................................. 6
  1.5. THESIS CONTRIBUTIONS ........................................................................................... 8
  1.6. KEY TERMS AND DEFINITIONS .................................................................................. 9

CHAPTER 2  SYSTEM QUALITY ............................................................................................... 11

  2.1. AN INTRODUCTION TO SYSTEM QUALITY ................................................................. 12
  2.2. THE IMPORTANCE OF SYSTEM QUALITY .................................................................. 13
  2.3. CONTRIBUTIONS TO SYSTEM QUALITY .................................................................... 14
  2.4. PRODUCT QUALITY .................................................................................................... 15
      2.4.1. Product Quality Dimensions ............................................................................. 18
      2.4.2. Trade-offs and Product Quality ........................................................................ 21
  2.5. PRODUCT QUALITY AND THE DEVELOPMENT PROCESS ...................................... 22
  2.6. EVALUATION OF PRODUCT QUALITY ...................................................................... 24
      2.6.1. Evaluation throughout the Lifecycle ................................................................. 25
      2.6.2. Evaluation of Internal and External Quality ..................................................... 27
      2.6.3. Evaluation of Quality in Use ............................................................................ 27
  2.7. QUALITY IN USE: A CONTEMPORARY VIEW ........................................................... 32
  2.8. BEYOND PERFORMANCE AND PRODUCTIVITY .................................................... 33
  2.9. CHAPTER SUMMARY ................................................................................................. 35

CHAPTER 3  AFFECT .................................................................................................................. 37

  3.1. AFFECT AND EMOTION ............................................................................................... 38
      3.1.1. Defining Affect .................................................................................................. 39
      3.1.2. Defining Emotion .............................................................................................. 41
      3.1.3. Clarifying the Differences: Affect and Emotion ................................................. 42
      3.1.4. Benefits of Distinction ..................................................................................... 44
      3.1.5. Terms Related to Affect and Emotion ............................................................... 44

INDEX ........................................................................................................................................ vii
6.3.6. Conclusions.............................................................................................................. 125
6.4. Chapter Summary.................................................................................................... 129

CHAPTER 7 VALIDATION OF AFFECTIVE REQUIREMENTS .............................................. 130
7.1. Stimulus-Affect-Measurement .............................................................................. 131
7.2. Quality in Use Evaluation Methods and Affect .................................................. 133
7.2.1. During-Use Evaluation Methods .................................................................. 135
7.2.2. Post-Use Evaluation Methods ...................................................................... 139
7.2.3. The Software Usability Measurement Inventory (SUMI) .............................. 142
7.3. Requirements for Affect Validation .................................................................... 143
7.4. Viable Affect Validation Methods ....................................................................... 144
7.4.1. User Self-Tracking ....................................................................................... 145
7.4.2. Psycho-physiology ..................................................................................... 148
7.4.3. Cued-Recall Debrief .................................................................................. 153
7.5. Summary............................................................................................................... 157

CHAPTER 8 EVALUATING USER AFFECT .......................................................................... 159
8.1. Aims for the Evaluation of Affect ...................................................................... 160
8.2. Pilot Study – User Self-Tracking ....................................................................... 161
8.2.1. Method .......................................................................................................... 162
8.2.2. Results .......................................................................................................... 163
8.2.3. Discussion and Outcomes of the Pilot Study .............................................. 163
8.2.4. Conclusion .................................................................................................... 164
8.3. Study – Psycho-physiology and Cued-Recall Debrief ...................................... 165
8.3.1. Purpose .......................................................................................................... 166
8.3.2. Method .......................................................................................................... 167
8.3.3. Pilot Study .................................................................................................... 182
8.3.4. Inter-Rater Reliability ................................................................................. 182
8.3.5. Results I Targeting Primary Study Aims ...................................................... 183
8.3.6. Results I Targeting Secondary Study Aims ................................................. 196
8.3.7. Summary of Results ..................................................................................... 198
8.3.8. Qualitative Review of Methods and Measures ........................................... 198
8.3.9. Conclusions ................................................................................................. 204
8.4. Chapter Summary ............................................................................................... 207

CHAPTER 9 REVIEW AND DISCUSSION OF THESIS ......................................................... 209
9.1. Review of Research Objectives ......................................................................... 209
9.2. Future Research Directions ............................................................................... 213
9.4. Final Comment ................................................................................................... 216

REFERENCES ................................................................................................................ 218

APPENDIX A: QUALITY AND QUALITY STANDARDS SUMMARY ..................................... 232
APPENDIX B: DESCRIPTIONS OF THE HUMAN-CENTRED DESIGN PROCESS ........ 235

APPENDIX C: TESTING OF THEORIES RELATING AFFECT AND DESIGN ........ 239
C.1 Web Based Survey .................................................................................. 239
C.2 Initial Data Analysis .................................................................................. 245

APPENDIX D: PARTICIPANT SELF TRACKING ........................................... 248
D.1 Option 1 Sample Scale ............................................................................ 248
D.2 Option 2 Sample Scale ............................................................................ 248
D.3 Option 3 Sample Scale ............................................................................ 249
D.4 Option 4 Sample Scale ............................................................................ 250
D.5 Avatar ‘Short List’ .................................................................................... 252

APPENDIX E: PSYCHO-PHYSIOLOGY AND CUED-RECALL STUDY ............. 253
E.1 Participant Information Forms ................................................................. 253
E.2 Participant Consent Form ......................................................................... 255
E.3 Participant Video Release Forms .............................................................. 256
E.4 Participant Demographic Questionnaire .................................................. 257
E.5 Participant Demographics ....................................................................... 258
E.6 Inter-rater Reliability Forms .................................................................... 260
E.7 Inter-rater Reliability Assessments .......................................................... 281
E.8 Inter-rater Reliability Results ................................................................. 284

LIST OF PUBLICATIONS DURING THESIS PROCESS .................................. 285
List of Figures

FIGURE 1.1: THESIS ORGANISATION CHART ........................................................................... 7
FIGURE 2.1: AREAS OF QUALITY FROM AN ORGANISATIONAL PERSPECTIVE ........................... 14
FIGURE 2.2: McCall’s QUALITY TRIANGLE ............................................................................. 16
FIGURE 2.3: SOFTWARE SYSTEM QUALITY CHARACTERISTICS ................................................ 17
FIGURE 2.4: QUALITY IN USE ................................................................................................. 18
FIGURE 2.5: RELATIONSHIP BETWEEN PRODUCT QUALITY DIMENSIONS ............................ 19
FIGURE 2.6: QUALITY DIMENSIONS WITH OTHER INFLUENCING FACTORS .............................. 20
FIGURE 2.7: ISO 13407 SYSTEM LIFECYCLE PROCESSES ...................................................... 24
FIGURE 2.8: PRODUCT QUALITY WITHIN THE SYSTEM LIFECYCLE ........................................... 26
FIGURE 2.9: PRODUCT QUALITY CONTEXTS, DIMENSIONS, OUTPUTS ........................................ 28
FIGURE 2.10: TWO INTERLOCKS WITH DECREASED UsABILITY FOR INCREASED SAFETY .............. 30
FIGURE 3.1: VALENCE-AROUSAL DIMENSIONS OF AFFECT AND EMOTION .............................. 46
FIGURE 3.2: THEORETICAL REPRESENTATION OF AFFECT AND EMOTION OF A USER DURING SYSTEM USE . 48
FIGURE 3.3: AFFECTIVE HCI RELATED RESEARCH DIVISIONS .................................................. 54
FIGURE 3.4: STEREOTYPICAL SHUT DOWN OR POWER OFF BUTTON ON MOST COMPUTERS ........ 59
FIGURE 3.5: QUALITY IN USE DUALITY – PERFORMANCE AND EXPERIENTIAL ......................... 64
FIGURE 4.1: HISTORICAL ETYMOLOGY TIMELINE OF SATISFACTION .......................................... 76
FIGURE 4.2: CONTEXTUAL MAPPING OF THE SENSES OF SATISFACTION .................................... 78
FIGURE 4.3: COMPARISON OF GROUPED MEANINGS AND SENSES OF SATISFACTION .................. 79
FIGURE 5.1: DIMENSIONS OF USER-EXPERIENCE IN DESIGN ................................................... 87
FIGURE 5.2: SCHEMATIC MODEL OF THE UNIVERSE OF USER-EXPERIENCE ............................... 89
FIGURE 6.1: CORRESPONDENCE BETWEEN DESIGN AND AFFECT THEORIES .......................... 102
FIGURE 6.2: AGE GROUP OF PARTICIPANTS .............................................................................. 105
FIGURE 6.3: GENDER OF PARTICIPANTS ..................................................................................... 106
FIGURE 6.4: HOW PARTICIPANTS HEARD ABOUT THE SURVEY ............................................... 107
FIGURE 6.5: INITIAL SOLUTION SCREE PLOT ............................................................................. 115
FIGURE 6.6: COMPUTER GAME GENRES AND RELATIVE SKILL ORIENTATION ............................ 119
FIGURE 6.7: LEARNABILITY BY GENRE ...................................................................................... 121
FIGURE 7.1: QUALITY ATTRIBUTE SCENARIO .......................................................................... 131
FIGURE 7.2: AFFECT QUALITY ATTRIBUTE SCENARIO .............................................................. 132
FIGURE 7.3: AFFECTIVE QUALITY ATTRIBUTE SCENARIO FOR THE THINK-ALOUD EVALUATION . 135
FIGURE 7.4: AFFECTIVE QUALITY ATTRIBUTE SCENARIO FOR DIRECT OBSERVATION ............. 136
List of Tables

TABLE 2.1: USABILITY FACTORS, DEFINITION, AND ISO SUGGESTED MEASURES ........................................... 31
TABLE 3.1: DEVELOPMENT OF EMOTION – THREE COMPETING THEORIES .................................................. 42
TABLE 3.2: RESEARCH WITHIN AFFECTIVE HCI .................................................................................................. 52
TABLE 6.1: ORIGINAL SURVEY ELEMENTS AND THEORETICAL BASES ..................................................... 108
TABLE 6.2: SURVEY ELEMENTS, THEIR THEORETICAL BASES, AND RELATED SURVEY QUESTIONS .......... 110
TABLE 6.3: FINAL ROTATED COMPONENT MATRIX ............................................................................................. 115
TABLE 6.4: RESULTING FACTORS FROM THE PRINCIPAL COMPONENTS ANALYSIS ..................................... 116
TABLE 6.5: RESULTS OF STANDARD LINEAR REGRESSION .................................................................................... 117
TABLE 6.6: POST-HOC RESULTS FOR GENRE DIFFERENCES ........................................................................... 120
TABLE 6.7: DESIGN GUIDELINES TO ENGENDER ENJOYMENT ......................................................................... 126
TABLE 7.1: ELEMENTS OF A GENERIC QUALITY ATTRIBUTE SCENARIO ............................................................. 132
TABLE 7.2: PARTS OF THE AFFECT QUALITY ATTRIBUTE SCENARIO ................................................................. 133
TABLE 7.3: USER SELF-TRACKING AND VALIDATION REQUIREMENTS ............................................................... 147
TABLE 7.4: THEORIES RELATING AFFECT, EMOTION, AND PHYSIOLOGY ...................................................... 149
TABLE 7.5: PSYCHO-PHYSIOLOGY AND VALIDATION REQUIREMENTS ............................................................... 152
TABLE 7.6: CUED-RECALL DEBRIEF AND VALIDATION REQUIREMENTS ......................................................... 156
TABLE 8.1: AFFECT VALIDATION METHODS AND IDENTIFICATION OR INDICATION OF AFFECT .................... 160
TABLE 8.2: SUMMARY OF SELF-TRACKING OBSERVATIONS AND PARTICIPANT COMMENTS ....................... 163
TABLE 8.3: PARTICIPANT DATA .......................................................................................................................... 168
TABLE 8.4: EQUIPMENT USED FOR PHYSIOLOGY DATA COLLECTION .............................................................. 179
TABLE 8.5: VIDEO RECORDING DETAILS, GAME PLAYING SESSION ................................................................. 181
TABLE 8.6: AUDIO RECORDING DETAILS, GAME PLAYING SESSION ................................................................. 181
TABLE 8.7: AUDIO RECORDING DETAILS, POST-CUED-RECALL SESSION ..................................................... 181
TABLE 8.8: AVERAGES SPANOVA ....................................................................................................................... 184
TABLE 8.9: VARIANCE SPANOVA ....................................................................................................................... 187
TABLE 8.10: ABSOLUTE CHANGE SPANOVA ...................................................................................................... 189
TABLE 8.11: AVERAGES SPANOVA ..................................................................................................................... 193
TABLE 8.12: VARIANCE ANALYSIS .................................................................................................................... 193
TABLE 8.13: ABSOLUTE CHANGE ANALYSIS .................................................................................................... 195
TABLE A.1: QUALITY AREAS AND SUB-AREAS .................................................................................................. 232
TABLE A.2: SUMMARY OF SOFTWARE QUALITY STANDARDS – GUIDELINES ............................................. 233
TABLE A.3: SUMMARY OF SOFTWARE QUALITY STANDARDS – SPECIFICATIONS ........................................ 234
TABLED.1: COMPLETE PARTICIPANT DEMOGRAPHICS FOR PHYSIOLOGY STUDY .................................... 258
Chapter 1  Introduction

Quality is a vague term and means different things to different people depending upon their own experiences, needs, wants, and expectations. Like beauty, it can be said to be in the eye of the beholder. Individuals can decide whether they find a car nice to drive, decide whether they like a particular style of cuisine, or whether they found a product useful or enjoyable. However, simply because one person perceived there to be sufficient quality does not mean that the next person will. This is, of course, obvious, and is the reason why there are a range of automobiles to choose, why different restaurants provide different cuisines, and why people can – most times – choose which product they wish to use.

While individuals know what their personal quality requirements are and can assess these themselves, developers of products are not as fortunate – they do not necessarily know exactly what an individual wants or desires. A car manufacturer must understand what qualities people want in their vehicle, a chef must know the client’s palate, and a developer must know what a user wants within a product. Therefore, while individuals can easily determine what quality means to them, quality and what it means within the context of a specific product remains a vague concept to a developer.

System quality, as a field, provides a means to describe how well a system is ‘fit for purpose’ (Juran, 1979). System quality can be further described as the ability of a product to meet the stated and implied needs and wants of the user. The study of system quality requires methods to extract user requirements – the user’s needs and wants – and techniques to validate whether these needs and wants are being fulfilled. A system is considered to be of good quality if it fulfils the user’s needs and wants and of sub-optimal quality if it does not.
Every product has quality – good or bad. Major characteristics that have been previously identified to improve quality include functionality, reliability, usability, efficiency, maintainability, and portability (Bevan, 1999). This list of quality attributes assumes that a product has been developed to fulfil some sort of utilitarian purpose. That is, if a product is quicker, safer, or can produce more, then the product will be of a higher quality.

However, productivity, performance, and safety do not explain or address issues that are raised by products that have the sole aim of providing a positive experience in a user – such as computer games that aim to create fun and enjoyment for a user. These experiences, and others such as fear, frustration, thrill, may be labelled ‘affective experiences’. Affect describes the discrete, conscious, subjective feeling that may have an influence upon a person’s overall emotion (see section 3.1.1 for the development and support of this definition). This leads to a distinction between product types. There are:

- Performance-based technologies where performance and/or productivity are the key determinants of product quality; and
- Non-performance-based technologies where affect and/or other general experiences are the key determinants of product quality.

The idea of fitness for purpose relates as much to making a product achieve the objectives of performance or productivity as it does to achieve the objectives of making a product fun or enjoyable to use. Non-performance-based technologies are not well considered within the current literature related to system quality, which has tended to focus on quality within performance-based technologies. However, as stated the concepts underlying system quality (i.e. fitness for purpose) imply that it should address and explain quality for any system, regardless if it is performance-based or non-performance-based.

1.1. The Resurgence of Interest in Affect and Product Design

The idea that developers should take note of affect and its implications on product quality is not new. Research dating to the late 1980’s suggested that positive affective experiences during product use can enhance learning, product uptake, and user enjoyment (e.g. Malone and Lepper, 1987; Carroll and Thomas, 1988). Further,
positive affective experiences can moderate negative experiences such as frustration. However, subsequent years yielded little work with respect to affect, and only recently has affect re-emerged as a desirable design characteristic. As Marcus (2004) writes, “there is a recent wave of research into affect, emotion, and design”.

Designing to enhance the affective experience has become increasingly popular, possibly due to the recent increases in choice that users have with respect to products they use or buy. As many competing products are equivalent in terms of functionality and performance, developers have begun to pay attention to the affect-related qualities of those products as a way to distinguish their product from its competitors. For example, modern mobile phones are all functionally similar – each can make and receive calls with relative ease, text messaging is enabled on each one, etc. For these functionally equivalent products, the ‘added bonuses’ targeting user affect ultimately influence the buyer’s decision (Lindgaard, 2001).

In a motivational article, Glass (1997) said, “If you’re still talking about ease of use then you’re behind. It is all about the joy of use. Ease of use has become a given – it’s assumed that your product will work.” While the assumptions that a product will work and that ease of use is guaranteed are arguable, Glass clearly makes the point that ‘joy’ and other affective user-experiences are important and should be considered in the development of products today. For many products, such as mobile phones, developers are targeting user affect to increase the quality of the product, despite not providing any utilitarian benefit to the user.

The targeted affective experience may be positive or negative. For example, computer games are designed to have the user experience fun (Draper, 1999), and haptic interfaces have been used to simulate the stressful situations in an operating theatre to assist in training surgeons (CSIRO, 2000). Another class of product – ubiquitous computers – attempt to create a null experience by having the product so integrated into the environment that users do not even notice it – which in essence means that the lack of an affective experience is the desired quality.

Unfortunately, there is limited research, theory, and practical guidance for developers of non-performance based technologies. It is possible that the resurgence of affect
research is because researchers have noticed this void in knowledge in relation to non-performance-based technologies. It is also possible that the resurgence of research interest is due to a saturation of performance-based materials, and therefore researchers are seeking a new niche. Or, as suggested previously, the resurgence of interest may be in response to the increase of development focus on the affect-related qualities of a product. Regardless of the reason, affect and its influence on product quality are important to consider.

1.2. Impacts of Affect on Quality in Use

There are three aspects of product quality – internal quality, external quality, and quality in use (Bevan, 1999). Internal quality deals with the hardware, programming style, etc. External quality deals with the actual product properties, characteristics, and behaviours. Quality in use (QIU) is the user’s perception of the ability of the product to meet their needs and wants. According to Bevan (1999), QIU is the prime validator of overall quality. Internal and external qualities can influence QIU, but are not themselves indicative of overall product quality (this is further explained in section 2.4.1).

Affective experiences are able to alter judgements, expectations, and memory (Castro et al., 1998), meaning that affective experiences can alter a user’s perception of quality (i.e. QIU). Further, affect is the subjective feeling that may arise within a user during product interaction, and is not directly related to internal or external qualities. As such, affect can only be validated as part of QIU. The internal and external qualities may influence the affective experience, but it is the interaction between the user and the product that creates the actual affective experience.

The following narrative highlights the impact that affect can have on QIU:

Imagine that you are developing an educational program to teach geography. This program must be able to identify countries, their capitals and their flags, provide key characteristics of each of these countries, and also have an option to explore a country in more detail (typography, geology of the land, people and culture, languages). This educational program would be quite powerful. Yet if these functional requirements are all that were addressed in the design it would be difficult to get many people interested in using it. It would be boring.
With educational software people must be motivated to use it. They must want to learn from it. If a person is internally motivated then they will both be more receptive to the information and they will learn more (Brown, 1988). One accepted method to encourage learning is to make it fun, engaging, or enjoyable – and this holds true for educational systems. In fact, one system type that has stemmed from this is edutainment. “Where in the World is Carmen Sandiego?” is one early edutainment program that recognised this need, and was successfully used around the world to assist in teaching geography. Many times, people used this program by choice simply because it was enjoyable! (from Bentley, Johnston, and von Baggo, 2002b)

The impact of affect on QIU is apparent within the narrative. If the product is boring to use then people will regard the product as low in quality. If it was made enjoyable or engaging – addressing affect-related qualities – then the QIU of the product will be greatly enhanced.

It is not sufficient to accept that affect is important to QIU – it is just as important to validate whether the required quality has been met. There are many QIU validation methods that can be used for performance-based qualities. However, there are no established methods to validate affect-related qualities.

1.3. Research Objectives

Affect is emerging as a primary determinant of quality in certain types of products (i.e. non-performance-based technologies), yet there is little or no guidance with respect to the development of products which aim to create user affect. Historical quality guidance offers little insight into the development of non-performance technologies, and even

This thesis has the primary objective of identifying the importance of affect with respect to quality in use, and to suggest how affect can be addressed and validated within the context of system design. To achieve the objective three research questions were developed. The first question forms the basis for the thesis. The second and third questions are more practical in nature as they address design characteristics that can enhance affect and methods to validate affective experiences.
The first question was:

1. **Does affect impact system quality? If so, how?**

This thesis investigates what is currently known about the relationship between affect and system quality and seeks to describe how affect impacts upon the quality perceived by a user.

The second question was:

2. **What product design characteristics can aid in the creation of affective experiences in the user?**

This question encouraged the exploration of products where affect is a known contributor to QIU. A theoretical study, a survey, and experimental studies performed as part of this thesis aim to identify characteristics of these products that promote positive affective experiences.

The third question was:

3. **What is an effective method to validate affect as part of quality in use?**

While it may be possible to identify product design characteristics that promote affective experiences in the user, merely having these characteristics within a product does not imply that the product will create an affective experience in a user. It is necessary to validate that the product does create the desired affective experiences. Experimental studies performed explore several possible methods to validate affect as part of QIU.

It should be noted that the area of affect, emotion, and system use has become an area of intense research scrutiny during the development of this thesis, and as such the thesis plan needed to evolve over time to account for new findings and directions. Regardless, the objectives remained consistent, and this ensured focus was retained throughout.

1.4. **Thesis Organisation**

As mentioned, the primary objective of this thesis is to identify the importance of affect with respect to quality in use, and to suggest how affect can be addressed and validated
within the context of system quality. The thesis can be divided into four areas, as shown in Figure 1.1.

<table>
<thead>
<tr>
<th>Area</th>
<th>Topics</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>System Quality</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Affect</td>
<td>3</td>
</tr>
<tr>
<td>Theoretical</td>
<td>Satisfaction and Affect</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>User Experience and Quality in Use</td>
<td>5</td>
</tr>
<tr>
<td>Practical</td>
<td>Affect and Design Characteristics</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Validating Affect</td>
<td>7</td>
</tr>
<tr>
<td>Review and Conclusions</td>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

**Figure 1.1: Thesis Organisation Chart**

**Background.** This area provides the reader with necessary background to the foundations of this thesis – namely system quality and affect. An overview of system quality is provided in Chapter 2, with emphasis on QIU. The theoretical foundations of affect are introduced in Chapter 3, which continues by summarising research relating affect and product design and how affect can influence product quality.

**Theory-based research.** This area explores several concepts that are related to or have been used synonymously with affect. It is necessary to explore these concepts because some are well addressed within system quality literature. Chapter 4 addresses satisfaction that, as a component of usability, is considered already part of product quality. A study is described, which investigates how the term satisfaction is used within the English language and serves to clarify the differences between affect and satisfaction. Within this study of satisfaction, the notion of the user-experience is raised, and this becomes the focus of Chapter 5, where it is argued that the user-experience and QIU are equivalent. A conceptual model of the user-experience/quality in use is proposed which accounts for different types of products (i.e. performance-based and non-performance-based technologies).
Practical-based research. The thesis then changes focus from how affect can influence system quality to practical issues related to addressing affect during the systems-development lifecycle in order to enhance product quality in non-performance-based technologies. Chapter 6 reports on a survey that had participants rate individual elements of three theories linking affect and design in the context of a computer game that they enjoy. The results of the survey supported the development of design guidelines that should enhance user affective experiences. Although these guidelines are beneficial themselves, it is necessary to be able to validate any requirement including those related to affect. To this end, Chapter 7 critically reviews several QIU evaluation methods regarding their ability to validate user affect during product use. Based on this review, a set of requirements for an affect validation method was developed. Several previously unexplored evaluation methods were studied and reported in Chapter 8. These included participant self-report of affective state, cued-recall methodology, and physiological measures. The latter two showed promise as affect validation methods.

Review and Conclusions. Chapter 9 concludes the thesis. It reviews the thesis as a whole, and discusses the broader implications of affect within the context of system quality. It also considers future research that will extend upon work presented within this thesis.

1.5. Thesis Contributions

This thesis focuses on the issue of how to assess or compare systems when affect is a prime determinant of quality, and contributes in several ways to this area of research within human-computer interaction.

The first contribution that this thesis makes is an inclusive literature review regarding what affect is, and how it has been used within related literature. Further to this, a linguistic study was undertaken to determine how affect and seemingly synonymous words within the literature (e.g. satisfaction) differ in actual ‘everyday’ language use. This review and follow-on study provides a clear, concise understanding of what affect is in relation to system use, and provides the foundation for further study into this area.
This thesis continues by studying what system properties are related to affective experiences during system use. To explore this, several theories that seemingly address this issue are studied, and an integrated set of system properties that can contribute to affect during system use is identified.

The third major contribution of this thesis is the proposal of methods that could be used to validate user affect within the context of quality in use. Multiple experimental studies were performed to explore a variety of methods. Ultimately, two methods were identified as having strong potential for the validation of user affect as part of quality in use – psychophysiological measurement and cued-recall debrief. This thesis contains advice on applying these techniques for this purpose.
1.6. **Key Terms and Definitions**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>A short term, discrete, conscious, non-reflective, subjective feeling that may have an influence upon a person’s overall emotion.</td>
</tr>
<tr>
<td>Emotion</td>
<td>The person’s overall feeling. There are many contributing factors to an emotion, including (but not limited to) context, past experience, recent experiences, personality, affect, and the net cognitive interpretation of these influences.</td>
</tr>
<tr>
<td>System Quality</td>
<td>The ability of a system to achieve stated and implied needs of users.</td>
</tr>
<tr>
<td>System</td>
<td>A group of interacting, interrelated, or interdependent elements forming a complex whole. Within the context of this thesis these elements include the user, the products that the user interacts with (including hardware, software), the environment (physical and psychosocial), and context of use.</td>
</tr>
<tr>
<td>Non-performance-based</td>
<td>These are technologies where performance has little or no effect on the perceived quality. Other factors, such as affective experiences, are the primary determinants of quality in use.</td>
</tr>
<tr>
<td>based technologies</td>
<td></td>
</tr>
<tr>
<td>Performance-based</td>
<td>These are technologies where performance (e.g. efficiency, productivity) is the primary determinant of product quality. Common qualities of performance-based technologies include functionality, reliability, usability, efficiency, maintainability, and portability.</td>
</tr>
<tr>
<td>technologies</td>
<td></td>
</tr>
<tr>
<td>Product</td>
<td>Within this thesis a product is a technological ‘object’ which the user can interact with to achieve some goal or fulfil a desire. The product is part of the overall system.</td>
</tr>
<tr>
<td>Quality in use</td>
<td>The user’s perception of the abilities of the product to meet their needs and wants.</td>
</tr>
<tr>
<td>(QIU)</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2  System Quality

Quality -- you know what it is, yet you don’t know what it is. But that’s self-contradictory. But some things are better than others, that is, they have more quality. But when you try to say what the quality is, apart from the things that have it, it all goes poof! There’s nothing to talk about. But if you can’t say what quality is, how do you know what it is, or how do you know that it even exists?

Pirsig, 1984
(concluding paragraph, Chapter 15)

This quote encapsulates many of the issues that are faced when trying to analyse and determine quality of a system. Each individual has his/her own perceptions of what creates or what is quality – what is one person’s junk may be another person’s treasure. However, when designing and developing interactive systems it is not possible to target each individual’s perception of quality, rather it is the task of the designer or team of designers to understand what the majority of the population will perceive as quality within the overall context of design. If a large proportion of the targeted population considers the system to meet their needs and desires, then the product can be considered to possess sufficient quality.

Designers are not without support. The domains of quality assurance and system quality provide methods, guidance, and suggestions to help designers understand what potential users will perceive as quality and how to achieve a quality product. This chapter provides an overview of system quality literature and knowledge. It concludes by noting some of the areas where system quality literature is lacking in the context of modern systems (sections 2.7 and 2.8).
2.1. An Introduction to System Quality

System quality, in its shortest, simplest, and perhaps most accurate form, can be considered ‘fitness for purpose’ (Juran, 1979). A system with high quality is one that meets its intended purpose without limitation. For non-interactive systems (i.e. automated systems) this is simply whether the system can do a particular task reliably. For an interactive system (i.e. requires human interaction) it means that a user is able to achieve their goals or objectives without any obstacles being met. The definition ‘fitness for purpose’ appears straightforward and simple to apply, yet many obstacles arise when it is applied to interactive systems.

The first obstacle is defining purpose. The purpose depends upon the perspective and goals of a user. For instance, in an office setting, an administrator and receptionist will probably have a very different purpose for using a human-resources database. The difficulty of defining purpose is made even more difficult when considering the diverse range of potential users and contexts of use. There are the end-users (as in the office example), but there are also people who maintain the product, produce new features of the product, or who re-design the product to work on different platforms or in different contexts than the original design. To complicate matters further, people will often find novel and unforeseen purposes and contexts in which to use a product – meaning that it is impossible to predict the purpose in these situations.

The second obstacle is defining fitness. Within system quality, it is necessary to be able to define criteria that the users of the product will see as enhancing quality. The degrees to which these criteria are met reflect the fitness of the system. However, these criteria are dependent upon the specific user’s purpose. For example, it may be sufficient for one user to wait thirty seconds for a product to respond, for another five seconds is the maximum permissible time, and yet another user time may not even consider time important. Thus, the fitness of a system again depends upon the needs and desires of the user.

So, while ‘fitness for purpose’ seems straightforward and simple, there is an underlying complexity to it. There is an escalating number of purpose x fitness criteria combinations as one addresses unique user groups and their specific needs or desires.
To assess overall quality of a system is virtually impossible, as there are so many different contexts to consider.

There is guidance material that will aid in developing and assessing the quality of a product throughout its lifecycle (e.g. ISO 9126, 2000; Earthy, 2001). Methods and techniques are provided to help ensure that quality requirements are considered early and often throughout the development process. The guidance also provides methods to assess the quality of a system, with the output of the assessment identifying ways that a product could be further improved to meet the changing and increasing demands of users. It is this guidance offered by the system quality domain that provides value to a development team.

2.2. The Importance of System Quality

System quality guidance material provides techniques and methods that a development team can use to accurately predict the characteristics of a product that lead to a quality system, which can be defined as the characteristics of a system that allow it to satisfy stated and implied needs of a user (e.g. Juran, 1979; ISO 9126, 2000). Normally this results in a more efficient, effective, and more satisfied user.

From a business perspective, marketing and selling a quality product will be easier because the quality of the product will “sell itself”. Further, higher quality products will often have a longer shelf life because they will not be superseded as easily or as quickly by competitor products that capitalise on the sub-optimal quality of the original product (Cooper, 1979). Thus, in the marketplace a high quality product has a greater likelihood of success, where success can be measured by user uptake, sales, financial profit, etc. However, it is necessary to note that product quality and the success of the product are not equivalent. Advertising, marketing, the economy, and consumer bias (e.g. brand preference) can all influence a product’s success. Further, a product can sometimes become a success independent of its quality – such is the case when there are no competing products. Therefore, while system quality can provide an indicator of success it is not the only factor contributing to success. Regardless, the argument stands that a higher quality product has a greater chance of success.
In short, overall system quality will enhance the match between the user and product (creating a better system), increase the life of the developed product, and the product will be easier to ‘sell’.

2.3. Contributions to System Quality

To understand how a quality product is developed it is necessary to look at the overall development process, including the organisation that is involved in development. Figure 2.1 outlines three key areas that can influence quality – organisational capability, process quality, and product quality (which is comprised of internal and external quality, and QIU).

Figure 2.1: Areas of Quality from an Organisational Perspective
(adapted from Bevan, 2001)

Organisational capability reflects the ability of the organisation to use ‘best-practice’ methods during development, and is the basis upon which other aspects of quality are built. Organisational capability asks, “Can an organisation implement a development process that will address functional, quality, and user needs throughout the development lifecycle?” The answer is dependent upon resources available, such as time, money, experience, skill, and knowledge. Organisational capabilities and their influences on product quality extend beyond the scope of this thesis and are not explored further.

Process quality builds upon the organisational capabilities. It draws upon the experience, knowledge, and abilities of employees to use industrial best-practice development methods. In effect, process quality asks, “Does an organisation implement a development process that will address elements of product quality throughout the development lifecycle?” Process quality includes the organisation’s promotion and
support of these activities – whether or not the organisation actually devotes the resources necessary to achieve quality in their development processes. Process quality and the influences on product quality are briefly discussed further in section 2.4.

Internal quality and external quality are properties that result from the development process. Internal quality relates to the static properties of the product (i.e. software code and hardware used), and external quality deals with the behaviour of the product independent of the user. The final level of quality, and according to Bevan (2000) the design goal of any interactive product, is quality in use (QIU). QIU is considered the true validator of overall system quality. If an end-user can successfully achieve his/her goals without obstruction in a given context, then the system can be considered to be of sufficient quality for that user – regardless of the internal or external quality of the product. Product quality – the combination of internal quality, external quality, and QIU – is discussed in detail in section 2.4, with the emphasis being on QIU.

There are international standards available that provide detailed information and support to each of these areas of quality within an organisation. Standards supporting all levels of system quality – including organisational, development, and product quality – are summarised in Appendix A:.

2.4. Product Quality

While it is important to know what can contribute to the likelihood of a quality product, it does not assist in describing what quality actually is or means for a product.

McCall, Richards, and Walters (1977) described product quality in terms of the properties that contribute to the existence of quality within that product, with the outcome being “McCall’s Quality Triangle” (Figure 2.2). McCall et al. did not define product quality – rather, they identified elements that contributed to product quality and by doing so provided insight into what product quality actually is.
This triangle provided an early classification of characteristics that could contribute to quality for a given context of use. Factors that contributed to quality were grouped according to their relevance during product operation, product revision (maintenance and upgrades), or product transition (adaptation to a new environment). This is an important feature of McCall’s Quality Triangle as it recognises that different user groups will have different needs when interacting with a product, and that their perceptions of quality will be based upon different factors. Like beauty, quality “is in the eye of the beholder”. To supplement the Quality Triangle, McCall et al. (1977) also provided a “quality equation”. This allows each quality factor to be given an appropriate weighting, and after these qualities are measured (in some manner) the outcome would be a single ‘product quality’ value. This formula is shown below.

**Formula 2.1: Product Quality**

\[
\text{Product Quality} = \sum (c_i \times m_i)
\]

- \(c_i\) is a weighting for the \(i^{th}\) factor
- \(m_i\) is the metric for the \(i^{th}\) factor

**Product Quality**: the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs.

The ISO/IEC 9126 standard elaborates upon this by including characteristics and sub-characteristics that can contribute to product quality. These are shown in Figure 2.3.

![Software Quality Characteristics](image)

Figure 2.3: Software System Quality Characteristics

The ISO/IEC 9126 standard was revised in the year 2000, and this saw the grouping of characteristics and their relative importance to QIU for specific classes of user (Serco Usability, 2001). Using the same six categories of product quality as in the 1991 standard, the new groupings suggest that the end-user is most interested in the functionality, reliability, efficiency and usability of a product, whereas the support user is more concerned with the QIU of maintenance and portability aspects (see Figure 2.4).

The groupings in ISO/IEC 9126 (2000) are comparable with the groupings used in McCall’s Quality Triangle. That is, qualities important to an end-user correspond with the qualities within McCall’s ‘product operation’ group, and the qualities important to the support user correspond to the McCall’s ‘product transition’ and ‘product revision’ groupings.
While the revised ISO 9126 standard and McCall’s Quality Triangle are similar, their emphasis seems to differ. The Quality Triangle has a product-centric perspective identifying the characteristics of a product that will add quality for certain uses. ISO/IEC 9126 takes a user-centric perspective, focussing on what people need to experience to perceive quality while using product. This change from product-centric in 1977 to user-centric in 2001 is consistent with changes in design philosophy that have occurred during the same period. The aim of design is no longer simply to create a functional product for a given context; rather the product must also assist the user in achieving their needs and objectives in an effective, productive, satisfying, and safe way.

### 2.4.1. Product Quality Dimensions

Product quality is considered to have three distinct but related dimensions: internal quality, external quality, and QIU (Bevan, 1999). The relationship between internal,
external, and QIU dimensions can be shown as in Figure 2.5, and is discussed further in section d below.

![Diagram showing the relationship between product quality dimensions](image)

**Figure 2.5: Relationship between Product Quality Dimensions**

*From Bevan (1999)*

### a. Internal Quality

The internal quality is the ‘core’ of the product and corresponds to the static properties of the product (i.e. software code and product hardware). Internal quality is the base quality upon which the other dimensions of product quality depend. If a product begins with poor internal quality then the external quality will be negatively influenced, and subsequently this negative influence may subsequently be translated to QIU.

### b. External Quality

External quality deals with the behaviour of the product. Often, external quality is concerned with whether the product behaves and performs as it is expected to in an environment with no external intervening variables influencing the behaviour of a product.

### c. Quality In Use

QIU is defined as “the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety, and satisfaction in specified contexts of use” (ISO/IEC 9126-1, 1991 as cited in Bevan, 2001). It addresses the behaviour of the product in the defined context of use while being used by a targeted user. QIU can be considered the ‘ultimate goal’ for product quality because if it meets the needs of the users and other stakeholders then it will be regarded as a quality product for that context, regardless of the internal and external quality.
d. Relating the Product Quality Dimensions

High internal quality does not guarantee high external quality, and similarly high external quality does not guarantee high QIU. Instead, they can simply influence the quality of a higher dimension, as shown in Figure 2.5. These are not the only influences on quality. There are many external influences that can add or detract from quality for each dimension. For example, the user characteristics have a strong influence on the QIU experienced by a particular user. If the user were familiar with a certain type of interaction style (perhaps form-based input) then he or she would probably find a product that uses this interaction style to be of higher quality than someone who is unfamiliar or uncomfortable with that style. Other influencing factors for each product quality dimension are shown in Figure 2.6.

![Figure 2.6: Quality Dimensions with Other Influencing Factors](image)

This figure is an extension of Figure 2.5. It integrates many influences discussed by Bevan (1999 and 2001).

However unlikely, it is possible for a product with low internal or external quality to have a high QIU. Consideration can be made for past programming styles. A quality product (one that satisfies stated and implied needs) was achieved despite the internal quality of the product being poor by today’s programming standards, as the programming styles adopted used to be ‘whatever works’. Many of these older products are still in operation today, and are still considered to have sufficient QIU to continue using – an example being the systems used in many middle-aged nuclear control centres. Understandably, any significant changes, upgrades, or modification to safety-critical systems such as a nuclear control centre must meet very strict standards. As
such, most existing control centres continue to use the original products because they do work and are reliable – they have sufficient QIU. As the adage goes – “don’t fix it if it isn’t broken”.

This highlights another issue related to product quality. The ‘standard’ for what one would call a quality product is constantly changing and being redefined. Automobiles from the 1930’s would not perform well by today’s standards, yet they were considered to be of high quality when first introduced. Computers, televisions and radios are among other products where the first products would no longer be considered to be of high quality. The evolution of system quality is an intriguing issue and one that could lead to some interesting observations. However, this issue is beyond the scope of this thesis and is not addressed further.

2.4.2. Trade-offs and Product Quality

It is not possible to ensure that all quality elements exist within a product – nor is it necessary. Trade-offs must be made between the various requirements of the product for practical reasons.

There are finite resources (i.e. time and money) during the development process, and these resources must be allocated optimally to ensure that the most critical aspects of the product are developed. It is therefore necessary to prioritise the requirements, identifying which requirements are integral to the product, which ones need to be addressed to some degree, and which ones may add value but are not integral to the product (Boehm, 1988). Resources should be focused upon achieving the highest priority requirements, and if there are available resources following this then the other ‘value adding’ requirements could be addressed (Karlsson and Ryan, 1997).

It is also possible for requirements to be at direct odds with each other, and thus prioritisation of the requirements is necessary to identify which quality is the most important for the product. An example given by Bass et al. (2003) is that reliability and security are often in a state of mutual tension: the most secure system has the fewest points of failure (such as a security kernel), but the most reliable system has the most points of failure (typically redundant processes or processors) where the failure of any one will not cause the system to fail. So if the product is required to be secure and
reliable then there must be some compromise between these two elements, as it is doubtful that both can be entirely dealt with.

Stakeholder representatives – including users – should assist the development team with the prioritisation of requirements to ensure that the requirements that will add greatest value to product quality are identified, providing developers an understanding of where they should devote resources. Further, involving the various stakeholders during this process may highlight and resolve issues relating to conflicting requirements.

2.5. Product Quality and the Development Process

It has been argued that the human aspect is often considered “too little and too late” during product development, resulting in systems with sub-optimal quality (Bias and Mayhew, 1994). Frequently the human associated only with the user interface of a product, and not as a component of the overall system. However, the human is the most complex and difficult component to design for in an interactive system (Bailey, 1996), and the human’s needs, wants, and capabilities must be considered from the outset. If not, the resulting system may not optimally divide task demands between the human and the product. Worse still, the resulting product may be based on invalid assumptions about the user and may make it impossible for the user to achieve his/her desired goals through product use.

With reference to the development process, system quality guidance material encourages the use of human- or user-centred design methods which:

- Promotes an enhanced understanding of the user;
- Promotes an increased understanding of how the product is to support the user in achieving his/her goals;
- Provides methods that enable the assessment of the developing product to ensure that the needs of the user are being addressed at each stage of development; and
- Provides developers the opportunity to recognise and respond to newly identified or changed requirements.
Human- or user-centred design methods place the human as a central figure in the development process, and aim to optimise the fit between the human and product. These methods require the human aspect of the system to be considered early and often (as opposed to too little too late).

A stereotypical example of a human-centred design process is provided by ISO 13407: Human-centred design processes for interactive systems (ISO, 1999 as cited in European MultiMedia Usability Services, 1999). ISO 13407 identifies four core principles that are necessary to achieve a human-centred design process. These are:

- An active involvement of users and a clear understanding of user and task requirements;
- An appropriate allocation of function between user and technology;
- Iteration of design solutions; and
- Multi-disciplinary design team.

Figure 2.7 shows the stages of a human-centred development process (each stage of this lifecycle is elaborated on in Appendix C: ). Human-centred development is an iterative process, where information gained from previous stages and previous iterations can be used to improve understanding of the contexts of user and user requirements, and enhance the design. At the core of human-centred development processes is the involvement of users throughout. This process should continue until the evaluation shows that the all requirements are sufficiently met, indicating that the product is fit for purpose.
It is important to consider the quality of a product throughout the iterative stages of the development process. Any quality deficiencies identified early in the development may have a large impact on later stages of development. Neglecting quality until later in the development process often results in the need to redesign and integrate the ‘missing’ quality attributes that will make the product viable for use in practice – and not because of functional deficiencies (Bass, Clements, and Kazman, 2003). The costs and time associated with redesign are considerable, and these can be avoided when the overall system quality is addressed early in development. By identifying and fixing quality deficiencies early there is an improved likelihood that the final system will be of high quality and at a reduced cost. Further, if a product is deemed to be of high quality then the system will have a greater likelihood of being successful in the marketplace than a product considered low in quality. How product quality can be tested throughout the system development lifecycle is discussed in section 2.6.

2.6. Evaluation of Product Quality

Understanding and developing for internal, external, or QIU will enhance the likelihood of a high quality product. However, understanding and developing for quality does not guarantee quality. Simply saying that a product was designed with quality in mind is inadequate proof that the product actually possesses quality. Similarly, following a human-centred design process does not guarantee quality. Without evaluating the product it is impossible to judge whether or not the product is good – or in other words
‘of sufficient quality’. As discussed in section 2.4, quality should be considered throughout the development lifecycle to ensure that the product is meeting the requirements of the user and organisation. It is the role of the evaluation process to indicate whether these requirements are being met.

Evaluating the quality of a product is particularly important during the early stages of system development. The early identification of requirements that are not being met will allow developers an opportunity to address them, and therefore improving the likelihood of a higher quality product. Further, the early identification of ill-met requirements will minimise the costs and time associated with redesign (if necessary) (Boehm, 1988). It is sometimes adequate to perform quick, informal evaluations to get an immediate impression of the product’s quality. With that said, formal evaluations will help ensure that all identified qualities are addressed in a valid and reliable manner.

In addition to ensuring that identified quality requirements are being met, an evaluation may also identify other areas of the product that could or need to be improved. For example, a product may be performing at or above all expectations with respect to all identified qualities. However, for this same product, user feedback may indicate that even though they were able to perform well they were not entirely confident with the product and how it works – effectively identifying another necessary requirement of training or documentation. This could lead to the refinement of documentation or the creation of a training course to better explain how the product functions. If either of these were implemented it may increase the user’s understanding of the product, which would lead to improved QIU.

2.6.1. Evaluation throughout the Lifecycle

Figure 2.8 shows how product quality dimensions are incorporated and evaluated throughout the system lifecycle (recall section 2.4. It illustrates that the evaluation of quality can occur during each iteration and each stage of the lifecycle. During early iterations, this evaluation may be based upon the described behaviour of paper-based prototypes, and later iterations may involve near-complete or complete working systems.
Figure 2.8 suggests that it is that real world needs of users (these can also be considered QIU requirements) will determine the external quality requirements, and in the same way the external quality requirements determine the internal quality requirements. This means that the internal quality requirements are actually derived from QIU requirements. Conversely, internal quality characteristics will influence the characteristics of external quality that subsequently influence actual QIU. Therefore, the verification of internal quality requirements and validation of external quality requirements are both – to a degree – representative of QIU requirements. However, as described in section 2.4.1, there are other external influences on QIU requirements and therefore internal and external quality are not entirely representative of QIU.
2.6.2. Evaluation of Internal and External Quality

Internal product quality can be measured through inspection of the code and assessment of the materials used. Internal metrics for software can include correctness of the code (e.g. are there any errors or bugs), structure of the code (e.g. how well does it conform to a particular programming standard), or coding efficiency (e.g. are there excess steps or procedures that must implemented for the product to perform). For hardware it could include the durability of materials (e.g. will they last), or the attributes of the materials (e.g. will the materials perform under required conditions) (Bevan, 1997).

External product quality can be measured through observation of the dynamic properties of the product. Measures can address any of the quality characteristics that have been identified as important for that specific product and their overt impact upon the behaviour of the product. Some examples include the latencies during the response to input, and the actual validity of output. External quality is often considered a necessary precursor to QIU.

Both internal and external quality are important in a product to enhance the likelihood of high QIU. These qualities are not elaborated on further, as the remainder of this thesis focuses on the QIU of products only.

2.6.3. Evaluation of Quality in Use

Bevan has argued on many occasions that QIU is the one true measure of overall system quality for an interactive product, and that QIU can be considered the major aim of product development (see Bevan, 1997, Bevan, 1999, or Bevan, 2000). If an interactive product has high QIU then users will consider the system to be of high quality, regardless of the actual internal and external quality.

The evaluation of QIU should target the product’s effectiveness, productivity, safety, and satisfaction of the user as these have been identified as the desirable constructs of QIU (see Figure 2.4). Figure 2.9 summarises the relationship between internal quality, external quality, QIU, and the desirable outcomes of QIU.
This figure differs from the original by showing the relationship between the quality dimensions (from Figure 2.5) and also showing the outcomes of quality in use (integrating aspects from Figure 2.4).

It should be noted that the effectiveness, productivity, safety, and satisfaction would be based upon different requirements for different users (recall Figure 2.4). For instance, the end-user will usually perceive a product as effective, productive, safe and satisfying if it has suitable functionality, reliability, usability, and efficiency. By way of contrast, the support user’s perception of effectiveness, productivity, safety, and satisfaction will be based upon the maintainability and portability of the product. During QIU evaluation, only those quality requirements that are important for the targeted user(s) need to be included.

A complete QIU evaluation is achieved through multiple evaluation sessions. Each context of use and each targeted user group should be included to get a true representation of the overall QIU of the product. However, this is not possible in many development situations. Products are increasingly targeting a wider breadth of user-groups, and often it is only possible to include a sub-set of these groups during evaluation due to resource limitations. A typical word processor today, for example, is not only the domain of secretaries or report writers – they are being used to collect notes during meetings, for families to create greeting cards during the holiday season, and as a method to keep personal diaries. Therefore, while it is possible to identify and evaluate the key contexts of use, it is not possible to identify and evaluate all contexts of use simply because of the time and costs associated. As such, QIU evaluations for many
products will be based upon a constrained set of contexts and user groups rather than a truly representative sample. Of course, by addressing the key contexts of use and the primary targeted users it is possible to get a reasonable representation of QIU.

a. Quality in Use and Usability

Quality in use and usability share the aims of effectiveness, productivity/efficiency, and satisfaction, with QIU also having the aim of safety (see Bevan, 1995a for QIU measures, and ISO 9241-11, 1998 for usability measures). Further, both QIU and usability focus on the user being able complete his or her objectives or achieve their goals during product use. As such, QIU and usability can, for the most part, be considered synonymous (Bevan, 1995b).

According to Nielsen (2003) usability does not address the utility of a product. Rather, usability assumes that the product was designed for a purpose, and therefore there is some utility in it. An example of this is an Eskimo/Inuit with a freezer – while the freezer could be very usable, it would have no utility for the user (this lack of utility may or may not be expressed through user satisfaction – it would depend on if and how satisfaction was measured). QIU, on the other hand, does account for utility in a particular context of use. If the product does not assist the user in achieving his or her needs and desires, then the product would have low QIU. Thus, the Eskimo/Inuit’s freezer could have high usability and low QIU for that particular user and context.

The second difference is that QIU may also address safety in addition to the common aspects effectiveness, productivity, and satisfaction. Given that safety and usability are at times opposed to each other, the exclusion of safety in the context of usability does make sense. There are, for instance, products that have been designed to be less usable in order to promote safer work practices or activities. An example would be machine interlocks where users must first perform a ‘pre-task’ to make the product function (see Figure 2.10 for an example). While this pre-task can create inefficiencies resulting in decreased usability, the safety benefits can improve the overall QIU.
Vehicle Interlock: For a person to start a vehicle they must first take a breath check and only if their alcohol level is below the required limit are they allowed to start the car. (From Ministry of Transportation, 2002)

Machine Interlock: For a person to use a machine they must press both buttons. The buttons are designed so that they require two hands, and ensures that the user is unable to insert their hand into the machine. (From Standards Australia, 1996)

Figure 2.10: Two interlocks with decreased usability for increased safety

In practice, many (if not all) usability professionals extend their coverage of usability beyond the achievement of effectiveness, efficiency, and satisfaction. They will often assess the utility of a product for a given user group and context of use, going beyond Nielsen’s defined scope of usability. These professionals are addressing issues related to overall product quality rather than just usability. It must be remembered that usability is only one aspect of product quality (recall the discussion in section 2.4, and in particular Figure 2.4).

b. Measuring Usability as Quality in Use

Usability evaluation has a tradition of focussing on the end-user and what they need or want when interacting with a product. As usability and QIU share the majority of product output measures (effectiveness, efficiency/productivity, and satisfaction), it was reasonable for Bevan (1995b) to suggest the adoption of usability measurement methods as QIU measures. This allows QIU evaluation to draw upon extensive experience and refinement of measures, rather than developing similar measures independently. The three main factors contributing to usability – effectiveness, efficiency, and satisfaction – can be evaluated as described in Table 2.1.
Table 2.1: Usability Factors, Definition, and ISO Suggested Measures
(from ISO 9241-11, 1998)

<table>
<thead>
<tr>
<th>Usability Factor</th>
<th>Definition</th>
<th>Suggested Measures</th>
</tr>
</thead>
</table>
| Effectiveness    | Refers to whether or not the user can achieve their goals, sub-goals, or objectives, and the degree of accuracy which they can do this. | • Percentage of goals completed  
• Number of errors made |
| Efficiency       | Represents the amount of resources used to complete the task. Resources may include time, cost, physical effort, materials, or cost. It is often viewed as a combined measure of effectiveness with respect to resources put in. | • Time taken to complete task  
• Effectiveness / time |
| Satisfaction     | Assesses the comfort and acceptability of use of the system.               | • Employee absentee rate  
• Learnability |

Some measures suggested within ISO 9241-11 are contentious. For example – employee absentee rate as a measure of satisfaction is not widely used or accepted within a usability context. However, usability goals and measures described as part of ISO 9241-11 were developed for use in the context of an industrial or office setting. Indeed, ISO 9241-11 is labelled “Ergonomics requirements for office work with visual display terminals”, exemplifying the office and workplace orientation. This background has resulted in a set of goals and measures that adopt an office work perspective of usability, where productivity and performance are the main aims.

Satisfaction in the office context could be assessed through employee absenteeism, as it is known that if employees are not satisfied with their work context they do take an increased number of sick days (Kim et al., 2002). However, the use of absentee rate does not apply to many products used outside of the workplace. For example, absentee rate is irrelevant for an automatic-teller-machine. For such products, satisfaction is more appropriately evaluated as to whether the product met the user’s expectations, whether the users liked using the product on, and whether the user would use the product again.

A product would be considered highly usable if the users were able to complete their tasks effectively, efficiently, being satisfied with their performance. A product would be considered to have poor usability if it failed to achieve suitable levels of effectiveness, efficiency, and satisfaction. While QIU would also address safety issues.
concerning the product, for most products high usability will indicate high QIU within performance- and productivity-oriented contexts.

2.7. Quality in Use: A Contemporary View

QIU, usability, and system development standards tend to be oriented towards performance-based technologies that range from word-processing software to safety-critical systems. However, there are many examples of modern products that deviate from this context, where productivity and performance are not the major determinants of QIU. There are products that have been developed solely to create an experience to the user such as fun, enjoyment, or even stress. Some products that fit into the category of non-performance-based technologies include:

- Haptic Interfaces
- Computer Games
- Ubiquitous Computing
- Children’s Toys
- Mobile Devices for Personal Use
- Educational Software / ‘Edutainment’

To add to this, many activities that users undertake do not have tasks that can be assessed in terms of productivity or performance. For example, surfing the web for leisure, playing a game, or having fun with a ‘new toy’ are not performance or productivity related, rather they are done for the experience. Many of these non-performance-based technologies and activities are of discretionary-use. This differs from performance-based products where use is often necessitated by the need to achieve a goal.

This does not mean that current system quality standards, evaluation measures, or theories are incorrect or inadequate – rather it simply highlights that they have a context in which they can be applied, and that context is currently performance and productivity.
2.8. Beyond Performance and Productivity

So why do people choose to use these products or undertake these activities if there is not a utilitarian need? Presumably, the user wishes a particular experience associated with a particular product. As an example, consider Norman’s (2004) description of the experience he derives from three unique teapots that he owns. One teapot is impossible to use. The second works quite well. The third can be considered the best teapot for making tea as it caters to all the stages of tea making, however it is also the most arduous to use. Norman says he does not use any of them – he uses a regular teapot because it is the most efficient way to make a cup of tea. He describes his attachment to the three teapots as being an emotional attachment – they are simply something he likes to look at, something he derives a pleasurable experience from, just knowing they are there. This description by Norman indicates that performance and productivity are not the only factors that are important in the design of a product, and hints upon another aspect important to design – that of aesthetics.

Aesthetics within a design is certainly an important factor. Anatole France (an artist) stated “beauty is the greatest power in this world” (as cited in Karvonen, 2000). Karvonen continues by saying that if this is true then “only a fool would neglect trying to understand such might” in the context of design. However, within systems development, where the focus has been on productivity and performance, that is exactly what has happened.

There were early advocates who tried to encourage people to see beyond the performance and productivity aspects of design (e.g. Carroll and Thomas, 1988; Laurel, 1991), but their writings failed to make a significant impact within the development communities. Only recently has the development community begun to recognise the importance of non-performance qualities. While there are varying possibilities for the recent uptake of interest in the area of user-experience beyond performance and productivity, one perspective is that it is an evolution in the progress of design – originally there was a need to ensure that things were able to perform adequately, and once that need was sated it meant that people could turn to higher order needs like pleasure, aesthetics, affect, and emotion.
Green and Jordan (2002) have set the stage for the adoption of non-performance-based qualities as a valid design goal. They identify three paradigm shifts in systems development in order to design for ‘emotionally rich interactions’. These are:

- “Don’t think affordances, think temptation.”
- “Don’t think ease of use, think enjoyment of the experience.”
- “Don’t think beauty in appearance, think beauty in interaction.”

These statements highlight that current systems development practices are oriented towards performance-based design. Affordances and ease of use are directly related to performance, and beauty in appearance suggests that it is simply a layer on top of the product and not something integral to the overall interaction. The shift towards temptation, enjoyment, and beauty in interaction acknowledges factors beyond performance and productivity, and accepts that user affect and emotional experiences (such as enjoyment) can be valid requirements and aims in development.

Many products will derive quality from both performance- and non-performance-based qualities. For a given product there tends to be an appropriate mix between non-performance- and performance-based qualities, depending on the targeted users and context of use. As such, for non-performance-based technologies it is still necessary to consider, prioritise, and make trade-offs between all qualities, performance and non-performance alike. Similarly, for performance-based technologies it is still necessary to consider non-performance-based qualities as these may influence QIU.

In short, non-performance-based technologies pose a problem to the current understanding of QIU. There is a limited understanding of non-performance-based qualities, especially as there has not been a reason to address these until recently (Hassenzahl, Beu, and Burmester, 2001). Existing QIU elements, such as those in Figure 2.4, are not necessarily relevant to non-performance-based technologies and what the user requires of these technologies. There is a lack of guidance regarding product or interaction characteristics that can lead to positive experiences or feelings within users. Related, there is a lack of published material related the general user experience or feelings of a user with system quality. Both of these voids in the
literature need to be filled to address the unique demands that non-performance-based technologies pose in the context of system quality.

2.9. Chapter Summary

System quality has been described as ‘fitness for purpose’, or alternatively the degree to which a product meets the stated and implied needs of users and stakeholders within a particular context of use. System quality, as a field, provides a wealth of information and knowledge regarding methods to ensure that the requirements of all stakeholders are considered throughout the development process, and provides criteria by which one can evaluate the degree to which the product meets these needs.

This chapter provided an overview of system quality as it currently stands, and provides the foundation for the remainder of this thesis. Specifically, it:

- Reviewed areas that can contribute to system quality. This included organisational capabilities to develop systems with high quality and development processes that can lead to quality products.
- Noted that it is necessary to balance resources and focus during design to ensure that key requirements – including quality requirements – have been adequately addressed.
- Identified that product quality has three dimensions – internal quality, external quality, and QIU.
- Discussed the need to evaluate internal quality, external quality, and QIU. QIU was considered as the defining measure of product quality, with the evaluation of internal and external quality being done to provide recommendations for improvement that should ultimately enhance QIU.
- Established that usability evaluation methods can be used for QIU. This means that QIU is able to benefit from the user-oriented evaluation methods that have been well developed within the usability context.

In addition to the overview of system quality, the final sections of this chapter (2.7 and 2.8) highlighted artefacts of product quality that arise when considering non-performance-based technologies. Specifically, it was:
• Asserted that product quality and usability were developed to address performance-oriented contexts, and as such does not adequately address qualities that were not related to performance and productivity.

• Suggested that as quality in use focuses on the degree to which stated and implied needs of users are met, quality in use should also address needs related to experiences such as fun, enjoyment, and stress.

The remainder of this thesis focuses on these artefacts of product quality in the context of non-performance-based technologies. The following chapter (Chapter 3) provides background to the concepts of affect and emotion (including differentiating between these), shows how these relate to product design, and presents a case of why affect is and emotion is not critical to address during the development process. Chapter 4 through Chapter 6 extend this knowledge through studies into the relationship of product use and experience.

Chapter 7 and Chapter 8 then look to integrate this knowledge with the ability to reliably evaluate the degree to which products create the desired affective experiences in the user. While it is always useful to describe what a user actually needs, without the ability to assess the degree to which this need is met, there is little assistance to product developers. Thus, the evaluation of affective experiences is fundamental to the integration of non-performance-based technologies into the scope of system quality.
Chapter 3  Affect

Utility and usability are important, but without fun and pleasure, joy and excitement, and yes, anxiety and anger, fear and rage, our lives would be incomplete.

From Norman (2004)

System quality focuses on the development of products that will meet the needs and objectives of people using them. Chapter 2 provided background to system quality, and noted that the primary emphasis has been oriented towards the performance and productivity of products. Yet, in a meta-analysis of usability studies, Nielsen and Levy (1994) established that performance only explained 21% of the observed variance in a user’s preference – suggesting that factors other than performance need to be addressed when determining QIU. So what, then, can explain the remaining 79% of the variance? Whatever the explanation, the current orientation towards productivity and performance seems inadequate to explain all aspects influencing QIU.

This is not an issue for all systems, as user preference is not always an issue in the selection and use of a product. For example, safety-critical systems are chosen and implemented based primarily upon their performance-based qualities. User preference has little influence on the QIU experienced for these systems, because the user will normally understand that these systems must function well above all else. However, not all products are oriented towards performance, and factors beyond performance will influence the QIU experience by users of non-performance-based technologies.

People are affective and emotional creatures (Norman, 2004). They experience and associate feelings when interacting with others, objects, and situations. This extends to product use. Several authors have suggested that the affective and emotive experiences of a user during and following interaction with a product will influence the user’s overall experience (e.g. Norman, 2004; Jordan, 2002). Therefore, as QIU reflects the
user’s interactive experience, it can be said that the affective or emotional experiences of a user can influence QIU. Neglecting the impacts of affect and emotion means the neglect of possible key ingredients of quality as experienced by the user.

QIU as it relates to performance-based technologies is well documented (as described in Chapter 2). While there has been some work in relation to affective requirements and quality in non-performance-based technologies (e.g. Johnson, 1999; Draper, 1999), there is still a need for more. Further, the available works tend to be written in an inspirational tone, emphasising the need to look at and consider affective qualities rather than delving into and exploring specific issues. This leaves a substantial scope for research relating affect, quality, and design with the aim of providing guidance and support for designers of non-performance-based technologies.

This chapter first defines and clarifies the distinctions between affect, emotion, and related concepts. This is important because these terms are used interchangeably, even though there are some important and beneficial distinctions. The chapter then focuses on affect within the context of design, and summarises major areas of research related to affect and design. Subsequent chapters delve deeper into specific issues related to affect and design – Chapter 4 investigating the relationship between affect and satisfaction, Chapter 5 exploring the concept of the user experience (and how this relates to affect and QIU), and Chapter 6 reporting on a study that identifies design characteristics that may enhance the positive affective experience for a user of a product. Combined, these three chapters aim to provide a clear reason why it is necessary to address affect within the context of QIU, and how to address affect during design.

3.1. Affect and Emotion

There is a “tidal wave of interest in emotions” within the field of human-computer interaction (HCI) (Marcus, 2004). This recent surge of interest in affect and emotion has resulted in many publications discussing related issues. However, these publications have neglected to make a distinction between affect and emotion (often using them interchangeably), and have failed to provide a clear and concise understanding of what is meant by either of these terms. Even a special issue on ‘affect and computing’ (Hudlicka and McNeese, 2003) did not provide a concise definition of
affect or emotion; nor did it make any distinctions between these two terms. This special issue described the issues related to affective computing that focus on making computers able to respond to and interact with people on affective and emotional levels.

The terms “affect” and “emotion” are frequently used synonymously. For instance, Hayes-Roth et al. (1998) and Arnold (1960) use the terms interchangeably within the same articles. Other times these terms are used alongside, without making any distinction – Norman et al. (2003) write about ‘the system of affect and emotion’ and discuss ‘emotional (affective) systems’. This swapping and interchange of terms can lead to confusion, particularly when researchers are using the same terms in different contexts. It is essential for a research area in its infancy and still seeking acceptance by a community – as affect and emotion within HCI are – that there is a clear understanding of the issues, that the aims of the research are precise, and confusion is minimised. Using terms interchangeably and in different contexts can lead to doubt as to the legitimacy of the research area.

Tomkin reflects on similar confusion that existed within the field of psychology in the 1940’s when he writes, “The field of affect and emotion was in deep trouble and disrepute at that time” (Tomkin, 1981). He continues by saying that this was due to inconsistent representations of affect and emotion, and owing to the inconsistencies, measures for affect were considered unreliable and lacking in validity regardless of the supporting research. The current state within HCI seems to be heading towards a repetition of history.

Affect and emotion are identified as distinct but related concepts within psychology research. Adopting the distinctions made between affect and emotion in psychology can have an immediate impact within HCI in increasing the clarity and credibility of affect and emotion research as related to product quality. These beneficial impacts are described in section 3.1.4, following the definitions of affect (3.1.1) and emotion (3.1.2), and a summary of key distinguishing factors (3.1.3).

3.1.1. Defining Affect

The definition of affect used throughout this thesis integrates research and descriptions of affect offered by Scherer (1994), Russel (2003), and Tomkin (1981).
Definition 3.1: Affect

Affect: a short term, discrete, conscious, non-reflective, subjective feeling that may have an influence upon a person’s overall emotion.

Scherer (1994) was interested in how affect impacts behaviour. In particular, he investigated how affective experiences manifested themselves in conversation by studying expressions during speech episodes, such as laughter or exclamation (e.g. “wow!”). He termed these expressions ‘affect bursts’, which were defined as “[affect bursts] are very brief, discrete, non-verbal, expressions of affect in both face and voice as triggered by clearly identifiable events”. The term ‘affect bursts’ was used in contrast to the term ‘emotional outburst’, where a person partakes in an “extended torrent of emotional tidings”. His definition of affect bursts brings in a qualifier of duration that can be placed upon affect to distinguish it from emotion. Affect is a brief, discrete response, whereas emotion tends to persist over a longer period.

It should be noted that affect does not require cognitive processing. This is noted by Russel (2003) who provides a detailed definition of ‘core affect’: “A neurophysiological state that is consciously accessible as a simple, nonreflective feeling that is an integral blend of hedonic (pleasure--displeasure) and arousal (sleepy--activated) values”. Russel’s definition of core affect is consistent with Definition 3.1, although this definition of core affect does not relate affect and emotion. However, Russel does address state that core affect is at the core of all emotion-laden occurrent events, which does suggest that affect will influence emotion. The definition of core affect was not adopted in its entirety because of its technical nature, and within HCI, a less technical definition is suitable for the breadth and background of practitioners involved – particularly as most individuals involved in HCI (practitioners and researchers alike) do not concern themselves with neurophysiology. Hedonic (alternatively called “valence”) and arousal values, as mentioned in Russel’s definition, are addressed in 3.1.5.

At times, people discuss ‘raw emotion’ – emotion that is pure and has not been subjected to rational thought. It is reasonable, and consistent with the definitions of affect and emotion used in psychology research and within this thesis, to label ‘raw
emotion’ as affect. However, the use of the term ‘raw emotion’ instead of affect should be avoided, as it is easily confused with emotion itself.

3.1.2. Defining Emotion

The definition of emotion used throughout this thesis is from Plutchik (1980), though there are several equivalent definitions available (e.g. Bindra, 1970; Arnold, 1960; Lazarus, 1982).

**Definition 3.2: Emotion**

*Emotion:* The person’s overall feeling. It has a variety of influences, including (but not limited to) context, past experience, recent experiences, personality, affect, and the net cognitive interpretation of these influences.

It is not possible to provide a complete listing of influences on emotion, as emotion is influenced by everything a human experiences, with ‘affect’ being one possible influence.

Researchers agree that there is a cognitive element to emotion, which functions to interpret a variety of influences to create a particular emotion. However, there is disagreement regarding the role or roles that affect may have in the formation of emotion. Three of these views are described and illustrated in Table 3.1. For clarity – it is necessary for a stimulus to be perceived for an individual’s affect to change (Tomkin and McCarter, 1964), but this does not mean that the stimulus was cognitively processed, as necessary for emotion. The perception of a stimulus occurs prior to cognitive interpretation, and it is possible for people to react to a stimulus without cognitive input (Wickens, 1996). This cognitive processing element of emotion is one of the key features that distinguish emotion from affect.
Table 3.1: Development of Emotion in Three Competing Theories

These descriptions highlight the role of affect within each theory.

<table>
<thead>
<tr>
<th>Source</th>
<th>Role of Affect in the formation of Emotion</th>
<th>Diagrammatic Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lazarus (1982)</td>
<td>Believes that people are constantly processing information, and based upon all information processed (including recent affective experiences) the individual will adopt an appropriate emotion.</td>
<td>![Diagram]</td>
</tr>
<tr>
<td>Zajonc (1980),</td>
<td>Suggests that a change in affect triggers the individual to interpret the situation and all contributing influences. Based on this interpretation, the individual alters their emotion to reflect the affective change and the changes in the overall situation.</td>
<td>![Diagram]</td>
</tr>
<tr>
<td>Teasdale and Fogarty (1979)</td>
<td>Offer the view that cognitive processes and affect act in a reciprocal relationship in the development of human emotion. That is, if an individual’s affect changes then the cognitive processes follow, whereas if an individual changes their cognitive focus (i.e. they focus on a new topic or situation), then this may trigger a change in affect. It is the combined effect of cognitive processes and affective experience that contribute to emotion.</td>
<td>![Diagram]</td>
</tr>
</tbody>
</table>

Support for each of these theories is available. Regardless, there is a consensus that the formation of emotion involves the cognitive interpretation of all relevant influences – including an individual’s affect. As such, the definition by Plutchik (1980) (Definition 3.2) remains valid as it recognises the role of cognitive processing and sees affect as one factor that can influence emotion.

3.1.3. Clarifying the Differences: Affect and Emotion

One possible reason for the confusion between these terms is the use of the same terms when describing emotions and affective experiences. For example, “happy”, “scared”, and “frustrated” can refer to affect just as it can to emotion. To illustrate, an individual may be “scared” when a person jumps out from behind a corner (affect), but once they
cognitively interpret all influences and realise that it is a friend playing a joke they may develop the emotion of “joy”. However, “scared” may also be used when a person is walking alone through a dark and unfamiliar area, and in this situation “scared” is describing their emotion as it involves the cognitive interpretation of many influences. The cross-use of these terms to refer to both affective and emotional feelings blurs the distinction between affect and emotion, and without understanding the context of the situation being discussed, it is impossible to decipher what is meant by “scared” or “happy”.

Three main points must be kept in mind to distinguish affect and emotion:

**List 3.1: Key Points to Distinguish Affect and Emotion**

- Emotion requires the cognitive interpretation of the influences, whereas affect does not.
- Affect is of short duration, whereas emotion reflects a longer-term feeling.
- Emotions are formed by the cognitive assessment of all influencing factors, only one of which is affect. Therefore, affect may have little or no effect at times, whereas at other times affect may be the predominant consideration.

The distinctions made between emotion and affect are congruent with the framework of emotions proposed by Russel (1999 and 2003), which seeks to further reduce the confusion that has existed in psychology emotion research since the 1940’s. However, instead of clarifying the terms affect and emotion, within his framework these concepts are ‘put in their proper place as folk rather than scientific concepts’ – meaning that affect and emotion are used too broadly to have any meaning and should not be used within scientific writing. Russel instead advocates the adoption of new terms in place of affect and emotion. To draw parallels, ‘affect’ as used within this thesis matches Russel’s term ‘core affect’, and ‘emotion’ as used within this thesis is congruent to Russel’s term ‘emotional episode’.

While the adoption of new terms, as proposed by Russel, would reduce confusion, it also requires people to accept, learn, and make use of these new terms. While the scientific and psychology communities may embrace these new terms, it may hinder uptake by the HCI community and the wider population. As the HCI community is
composed of experts, researchers, practitioners, and other interested parties from a variety of disciplines and as the HCI community regularly interacts with the wider population the refinement of old terms is preferable because it is less apt to disenfranchise people outside the scientific community. Further, as affect and emotion are widely used in related HCI literature, it was assumed that the introduction of new and discarding the old terms would not be easily disseminated.

3.1.4. Benefits of Distinction

Tomkin (1981) stated that the field of affect and emotion research (within psychology) was in disrepute in the 1940’s because of the inconsistent use of the terms affect and emotion. As a result, affect and emotion research were perceived as unreliable and lacking in validity. In a similar manner, the synonymous and inconsistent use of “affect” and “emotion” is causing confusion within HCI. It is desirable to learn from these past experiences, and to avoid the same difficulties that the field of psychology experienced. The distinctions between affect and emotion made within this thesis and summarised in List 3.1 aim to make it clear that affect and emotion are related but not synonymous. The distinctions will reduce the ambiguity between these terms, and the unified operational definitions provided will help people discuss, publish, and reflect on research with confidence rather than uncertainty.

A second implication – more practical and more oriented towards the goals of this thesis – is that the definitions of affect and emotion also clarify what QIU evaluation methods actually measure. For example, questionnaires, focus groups, and interviews will often ask users about their overall feelings towards a product. As such, they are querying the user about their emotions following product use and do not target the user’s affective experiences. Commonly employed product evaluation methods do not appear to be sensitive to affect (this is explored further in Chapter 7).

3.1.5. Terms Related to Affect and Emotion

There are several terms that are often used in association with affect and emotion that should be defined. The first is mood, which is important to note because it is coupled to affect, and emotion. The second set is valence and arousal. These are adjectives that can be used to further define affect and emotion.
a. **Mood**

“Mood” has not yet made an appearance in HCI related research literature, but is used in similar contexts to that of emotion and affect. Because of this, it is prudent to understand what it is and how it relates to affect and emotion. Mood can be defined as:

**Definition 3.3: Mood**

*Mood*: The general state of a person and includes affective components, cognitive interpretations of contextual components, and emotional components.

*From Mayer, Salovert, Gromberg-Kaufman, and Blainey, 1991*

This definition of mood identifies it as being the general state of a person (the authors of originally used the term “organism” rather than person) and can be comprised of several affective components, several contextual components, and several emotional components. This general state is not to be confused with the person’s overall feelings (i.e. emotion). The current emotion being experienced by an individual may influence mood, however there are many other influences that must also be accounted for. Further, mood is much more persistent than emotion, with mood changing gradually over time where emotion can change relatively quickly depending upon the situation (Mayer et al., 1991).

At any given time, an individual can be said to be in ‘a mood state’. This is the disposition of a person to react with negative or positive emotions in a given situation (Castro et al., 1998). Evidence suggests that most humans tend to be in an overall positive mood state when they are at ‘base level’ (Castro et al., 1998). That is, if an individual is in a neutral situation (i.e. no positive or negative influences), there is a tendency for that individual to react positively to the situation. Interpreted, it means that people see the absence of any negative influences as a positive (Castro et al., 1998).

Mood state can influence how people interact with a product and the experiences they have. If a person is in a positive mood state, the individual will have a more positive reaction than if they were in a neutral or negative mood state. This may explain why people who react to a product positively one day react differently the next. While it is feasible to design interactions to elicit specific affective experiences (see work
summarised in section 3.3.2) and perhaps even to design products that can influence emotion, mood may be too persistent to influence. That is – because mood is such a persistent feeling it may not be easily changed through product interaction. With that said, it may be possible that affective experiences during product interaction and resulting emotional experiences will cause a user’s mood state to shift – either positively or negatively.

b. Valence and Arousal

Valence and arousal are adjectives that can be used in association with affect and emotion. Lang (1995) describes these as:

- Valence: concerned with the **positive or negative** dimension of affect or emotion (happy vs. sad, excited vs. scared, unconcerned vs. worried).
- Arousal: represents the **intensity** of the emotion or affective response (excited vs. calm, energised vs. lethargic).

These dimensions are shown in Figure 3.1 with some terms describing the type of emotion or affective response that may be characterised by each.

![Figure 3.1: Valence-Arousal Dimensions of Affect and Emotion](From Lang, 1995)

The primary method of communicating valence and arousal to others is through human behaviour (e.g. body language) and non-linguistic aspects of speech (e.g. fluctuations in voice pattern). Humans are adept at estimating these dimensions in people that they know. Once a persons’ normal behaviour, actions, speech characteristics, etc. are
known it becomes possible to make relatively accurate judgements about their current affective and/or emotional state. People who are unfamiliar with each other are still able to make some judgements, however the accuracy of these judgements will be considerably less. High arousal amplifies the overt representation of affective and/or emotional experiences. Therefore, it is easier for a person to make accurate judgements about another’s affective and/or emotional state during high levels of arousal.

### 3.2. Why Affect and Not Emotion?

One key distinguishing feature between affect and emotion is the duration and immediacy of occurrence. Affective experiences are of short duration and are an immediate subjective feeling in response to a stimulus. Emotions, on the other hand, are a longer-term phenomenon and can be influenced by a variety of factors – some extending beyond the immediate user-product context. Affect was chosen as the focus of this thesis because it is more sensitive to the impacts of specific interactive elements, whereas emotion tends is less sensitive to the specific elements.

To illustrate, Figure 3.2 represents the possible relative affective and emotional experience of a user during product interaction. During interaction, two extreme affective changes can be noted – one positive and one negative. It would be possible to match the times of these extreme deviations with what the user was doing; thereby matching what aspects of the interaction triggered these experiences. Emotion is less subject to variation because it is more persistent than affect and because of the range of influences acting upon it, which can moderate the influences of affect. Throughout the interaction, the user’s emotions remained positive but, unlike affect, there were no significant deviations and it would not be possible to associate any aspect of the system that enhanced or diminished the user’s emotional experience. Providing developers with information about aspects of the interaction that create positive and negative affect allows improvement of the negative aspects and care can be taken not to lessen the impact of the positive aspects (assuming it is the positive experiences that are desired).
It could be argued that emotion is perhaps more important than affect – after all the resulting emotion is what the user will ‘walk away with’ and will be what they tell friends about. While this is probably true, emotion does not easily feedback to the development process because it is not easy to identify what aspects of the interaction created the resulting emotion.

Affect is better able to ‘pin-point’ aspects of a system that are creating positive and negative experiences, and as such can provide direct input into product development. Emotion is subject to too many influences, and it is difficult – perhaps impossible – to determine the effects of product interaction on a user’s emotion. However, it is theoretically possible to associate affective experiences with characteristics of the system. It is on this premise – the ability of affect to provide input into product development – that affect and its relationship to QIU is the focus of this thesis.

### 3.3. Affect and Computing Research

There has been a large amount of recent research and literature published in relation to affect and emotion. This array of work belongs to one of two key themes:

- **Affective Computing** (product focused): Understanding how products can respond to and cater for differing affective and emotional states in humans including the projection of emotion from the product; and
- **Affective HCI** (user focused): Understanding how product design and interaction influence a user’s affect and emotion.
Each theme is described in more detail in sections 3.3.1 and 3.3.2 respectively. This thesis would be considered part of affective HCI as it focuses on the user’s experience and QIU.

3.3.1. Affective Computing

The idea of computers that can respond to user affective states and emotions is not new. HAL, the supercomputer in the movie 2001: A Space Odyssey (Kubrick, 1968), conversed fluently with the astronauts aboard the ship and responded to their emotions – querying the astronauts why they felt a certain way and trying to offer suggestions regarding how they could make themselves happier. That movie highlighted the benefits of using computers to respond to affect and emotion as HAL was able to alter its tone and approach to the astronauts to encourage them to continue with their tasks or to improve their performance.

The term “affective computing” was adopted by Picard in 1995 when she used it to describe computers that can recognise and react to human affect (Picard, 1995). She asserted that these computers could benefit users by appropriately reacting to their affective needs, thereby using the optimal ‘tone’ of feedback to best motivate users to continue using the product – similar to HAL. Further, Picard saw the possibility that these computers can aid researchers to better understand the processes by which affect and emotion are influenced and created. It should be noted that researchers in the field of affective computing tend to use affect and emotion interchangeably, and it is presumed that they do indeed try to address both affect and emotion (as defined within this thesis).

Hollnagel (2003) argues that computing – by the very nature of the term – cannot be affective and that the term ‘affective computing’ is an oxymoron. Computing is composed of algorithms and is bound by the rules of logic. However, human affect and emotion is often without a logical explanation. Picard (2003) refers to the character ‘Spock’ on Star Trek as the sentient representation of a computer – logical, precise,

---

1 This same movie also highlights the extremes of this ability – when systems begin to be able to read affect and emotion in humans so well that they can foresee their own demise and as a result attempt to prevent it. The moral question of giving machines the ability to interpret emotion is well beyond the scope of this thesis, and is best left to the philosophers, story tellers, and movie makers.
rational, and has difficulty in understanding human emotion. As Spock often says, “I see no rational reason behind [a decision]” (Nimoy, 1995). This epitomises the Hollnagel’s argument – a computer cannot be ‘affective’ when it is limited by its logic, just as Spock was not ‘affective’ given his reliance on logic. However, and despite his criticism of the term ‘affective computing’, Hollnagel (2003) does support the underlying research area. He can see the benefits of computers being able to sense, adapt to, and simulate affect and emotion and does support Picard’s vision of ‘affective computing’ despite his sentiments towards the label.

The MIT Media Laboratory Affective Computing Research Group headed by Rosalyn Picard has made inroads into developing products that can sense, react to, and respond with seeming affective qualities (MIT Media Laboratory, 2004). Recent work published by this group has outlined the development of products equipped with an ‘affect-support agent’ that can sense and react to user frustration through seemingly human behaviours including active listening, empathy, and sympathy (see Klein et al., 2002 and Picard and Klein, 2002). In their studies, the researchers frustrated the user by giving them a product that was difficult to use, and found that users who interacted with the affect-support agent continued to use the product long after users who did not have the affect-support agent available. The affect-support agent was developed only to sense frustration and does not yet respond to other affects or emotions. To do this the affect-support agent identifies interactive behaviour that is commonly associated with user frustration (e.g. putting more pressure on the mouse, making more ‘erratic’ cursor movements, and typing with more pressure).

The results from Klein et al. (2002) suggest that a product that can interact with a user on an affective and emotional level can enhance performance and productivity. This can also be understood by recognising that human-human communication represents an ideal form of interaction as words are not the only thing that convey meaning – tone, body language, etc. can all add value to communication between humans. Computers ignore this ‘value added’ communication and human-machine communication is simple by comparison. Affective computing aims to mimic human-human communication by giving the computer the ability to recognise and respond to affect and emotion. Though the term ‘affective computing’ has only recently been adopted, artificial intelligence research has been pursuing a similar goal since the 1940’s. True ‘artificial intelligence’
is defined by the ability to engage in human-like conversation to the degree that a human judge cannot determine if they are conversing with a human or machine (Turing, 1950).

Despite the early successes within affective computing there are still sceptics as to the value it will have in systems development. Picard (2003) and Hudlicka (2003) have each attempted to answer questions posed by these sceptics. One of the most pertinent questions asks, “Do we really need to bother with affect?” The concise response to this offered by Hudlicka (2003) was “not always”. Hudlicka continued by saying that there are products where interacting on an emotional and affective level with a user is unnecessary. Further, it is possible that a product that adapts to the users affective state could be detrimental to overall system performance in certain contexts because it may create an element of uncertainty. By contrast, some products could be enhanced by recognising and catering for user affect and emotion. Current research in affective computing is focused on hostile or negative affective experiences (e.g. frustration) with the aim of reducing the intensity or eliminating these experiences through computer aided mediation.

There is one additional worry about making computers seem like they can feel, and that is that people will start treating them as though they can. Users currently know that it is impossible to offend a computer – no matter what the user says, types, or inputs into the computer it will continue to function as designed (assuming it is not broken). However, if the computer appears to have feelings then people may be apprehensive during their interaction with it for fear of offending it.

### 3.3.2. Affective Human-Computer Interaction

Affective HCI is a term that Hudlicka used to describe any product that is designed with the intent to elicit an affective response in a user (Hudlicka, 2003). To contrast, affective HCI can be considered human centric (focused on human affect and emotion caused by a system), whereas affective computing is product centric (focused on the ability of a product to react to human affect or emotion). As with affective computing, researchers in the area of affective HCI tend to use affect and emotion synonymously.
Affective HCI researchers believe that an understanding of the relationship between affect and design will result in informed decisions about design choices to create a desired user affect. Further, this understanding will minimise the likelihood that a product design will create a wrong or undesired affective experience. Hudlicka (2003) provides two examples of products that produced undesired user affect – talking cars and talking elevators. Developers of these products incorrectly assumed that providing feedback to a user in a pleasant voice would enhance the user’s experience. The ‘pleasant voice’ backfired – users found the voice annoying and intrusive, and would have preferred no voice at all. Affective HCI researchers aim to provide guidance for developers, helping them understand how to design for affect and to understand the affective experiences the user desires in a particular context.

There has been an array of affective HCI research published in various areas, most of it quite recent. However, there has not been any attempt to integrate the many streams of affective HCI research into an overall framework or broad understanding of the relationship between affect and design. Instead, affective HCI researchers tend to focus on a specific affect and design. Table 3.2 summarises some research related to affective HCI.
Table 3.2: Research within Affective HCI

<table>
<thead>
<tr>
<th>Research Area</th>
<th>Applicable Reference(s)</th>
<th>Summary of Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beauty</td>
<td>Karvonen, 2000</td>
<td>Looked at beauty and how it relates to aesthetic design</td>
</tr>
<tr>
<td></td>
<td>Fujino, 1999</td>
<td>Studied Japanese architecture to illustrate how beauty and utility have effectively been used together to create the optimal design.</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>Lindgaard, 2001</td>
<td>Found that people may like a web site for its aesthetic appeal, despite poor usability</td>
</tr>
<tr>
<td></td>
<td>Whitten II, 2001</td>
<td>Described a measure for user-experience that included usefulness, usability and aesthetics</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>Hassenzahl et al. 2001</td>
<td>Used elements of hedonic theory to classify attributes of products that can create enjoyable interfaces.</td>
</tr>
<tr>
<td>Fun</td>
<td>Draper, 1999</td>
<td>Suggested that fun should be considered a valid software requirement</td>
</tr>
<tr>
<td></td>
<td>Kersten-Tsikalkina and Bekker, 2001</td>
<td>Studied the relationship between usability and fun of children’s products (toys), and found that usability is only one possible influence on fun.</td>
</tr>
<tr>
<td></td>
<td>Carroll et al. 1988</td>
<td>Indicated a need to study fun within HCI</td>
</tr>
</tbody>
</table>

There appear to be three main areas of affective HCI research. These include theoretical views of affect adopted from psychological theory (e.g. Hassenzahl et al., 2001), the study of affective experiences during product use (e.g. Kersten-Tsikalkina and Bekker, 2001), and the study of interface design and its influences on user perceptions (e.g. Lindgaard, 2001). These areas are related, and often researchers will adopt ideas and techniques from another area if it will be of benefit.

As mentioned, affective HCI researchers have tended to focus only on a specific topic area, and have not explored the more holistic concept of affective HCI. Figure 3.3 represents a holistic representation of affective HCI, showing the contributions of each affective HCI research area. Each research division is described in the following
subsections. Subsection d provides a summary of how each of these divisions can contribute to a better understanding of the holistic concept of affective HCI.

![Figure 3.3: Affective HCI Related Research Divisions](image)

**a. Theoretical Perspectives of Design and Affect**

Joy – “you don’t notice it, but you’re drawn to it” (Glass, 1997). Glass suggests that joy is an experience that one wishes to have. Viewing joy from this perspective is consistent with the principles of hedonic theory, where an individual’s well-being consists of the subjective evaluation of happiness and pleasure versus displeasure (Ryan and Deci, 2001). Hedonic theory is derived from Greek philosophers who believed in that only what is pleasant or has pleasant consequences are intrinsically good (Simpson and Weiner, 1989).

Hassenzahl et al. (2001) applied hedonic theory when they discussed joy in the context of product interaction. They suggested that the ‘hedonic qualities’ of an interface can increase the joy experienced during product use, which would also increase the user’s overall well-being. Hedonic qualities focus on the human’s desire for novelty/change and social power – such as innovation or exclusiveness (Hassenzahl et al., 2000). Hassenzahl et al. (2001) do provide one caveat – hedonic qualities within an interface
may be at the cost of utilitarian (performance-based) qualities. This is not surprising as it is known that there will often be trade-offs between competing requirements and it is the developer’s role to ascertain which qualities should be included and to what degree (recall section 2.4.2).

In a study, Hassenzahl et al. (2000) contrasted the effects of hedonic qualities and performance /productivity qualities on product appeal. The participants in the study were asked to “switch off a pump in an assumed industrial plant” using the interface provided. This activity was chosen because it was simple, realistic, and was a context that the participants could relate to. Results showed that the experience of a user is a combination performance, productivity, and hedonic factors. They also suggested that an improvement in any one of these qualities can offset the effects of deterioration in any of the other qualities.

However, the task chosen by Hassenzahl et al. (2000) was based in an industrial context and was performance- and productivity-based, and presumably users wanted to complete that task. Defined tasks do not necessarily represent the types of goals related to the use of non-performance-based technologies. In non-performance-based technologies the user’s goal is often the experience and not dependent on the outcome of a task.

Also in this work, Hassenzahl et al. (2000) identified that emotions (and presumably affect as no distinction is made) are a consequence of design and are not considered design goals. However, there are many non-performance-based technologies where the design goal is to create affective and emotional experiences in a user (refer back to 2.7). For these reasons, the work by Hassenzahl et al. (2000) is limited to performance-based technologies, and not necessarily representative of non-performance-based technologies. Regardless, finding that hedonic qualities have an impact on the QIU of performance-based products – where affect and emotion are not design goals – suggests that in non-performance-based products hedonic qualities may have an even greater impact on QIU.

A competing theory to hedonism is eudaimonism, which is based upon the concept of self-actualisation. Eudaimonism sees well-being developed by living in accordance
with ones daimon or true self (Ryan and Deci, 2001). Eudaimonic theory suggests that it is only when one acts and behaves within the scope of personal values and morals that well-being is improved, and that the achievement of desires does not necessarily improve well-being. In the context of user-product interaction, eudaimonic theory suggests that users should choose and use products that reflect their personal values and morals, regardless of the achievement of desires and goals, and by doing so the experience will be positive. According to eudaimonic theory, the ideal interactive product would involve the user performing tasks or activities that are in agreement with the user’s values and morals, or – in the least – the tasks and activities should not be counter to those values and morals.

Evidence suggests that both hedonism and eudaimonism can influence well-being and experience (Ryan and Deci, 2001). As such, there also needs to be research into eudaimonic qualities of a product, which would relate to the use of a product that is in line with a user’s values and morals. Unfortunately, with the breadth of cultures and religions using the same product it is difficult, perhaps impossible, to make any assumptions about the values and morals of all potential users except in a few, rare contexts where the users are very homogeneous. As such, eudaimonic theory has yet to make an appearance in affective HCI research.

The adoption and application of theories such as hedonism and eudaimonism to human-computer interaction is difficult. These theories have various intricacies that must be acknowledged in order to apply them appropriately to a HCI context. The work by Hassenzahl and his colleagues (2000 and 2001) represents an initial foray into this area by applying hedonic theory to product use. It is expected that other researchers will begin to explore various affect-related psychological theories (perhaps eudaimonism) and to study their roles in affective HCI.

b. Affective Experiences and Design

In 1988, Carroll and Thomas cautioned the HCI community that concepts such as ‘fun and use’ need to be studied, not neglected or assumed irrelevant. They believed that HCI would progress into the realm of overall user-experience and beyond utility and usability. While Carroll and Thomas only discussed fun, other affective experiences were implied.
Fun has effectively been used as a way to entice people to experience new situations and has been used as a teaching tool. “Where in the World is Carmen Sandiego?” (The Learning Company, 1985) is an early educational game that recognised fun as a method to entice children to learn – in this case geography. The designers of this game recognised that if people wanted to use the product (i.e. they were internally motivated) then they will be more receptive to the information and will subsequently learn more (Brown, 1988). By making the product fun schoolchildren wanted to learn and had fun doing so.

In 1998, the idea that fun is a factor in design began to re-appear in HCI research and literature, and in 1999 Andrew Monk chaired a workshop titled “Fun and Computers” (Monk, 1999). This workshop encouraged researchers and practitioners to provide input into:

- the perceived role of fun in computer supported cooperative work for children,
- the viability of user-centred design practiced in ‘entertainment computing’,
- cognitive models to explain fun with computers,
- fun and impacts upon memory, and evaluation of fun in design.

This workshop has been run two additional times (see Monk, 2000 and Monk, 2001), with contributions expanding into the fields of hedonism, interactivity, learning, overall user-experience, humour, children’s toys, and more generally affect and emotion.

Inspired by the first “Fun and Computers” workshop and the questions it implied, Draper (1999) took a software engineering perspective and was adamant that developers must consider fun as a valid software requirement. Draper also recognised the influence that motivation can have on product interaction and affect. He suggested that extrinsically motivated use (such as ‘work’ motivations) cannot create fun or enjoyment – only intrinsically motivated activities can do this. This is not to say that a product used in a ‘work motivated’ context cannot be fun or provide other positive affective experiences. It is possible for intrinsic and extrinsic motivators to co-exist (Malone and Lepper, 1987), and while ‘work’ motivations are normally extrinsic, the user may have intrinsic motivations to use such a product. For example, a worker may be required to use a product to perform their normal duties in order to be paid (extrinsic motivation).
At the same time, the same worker may try to work faster to finish early on a particular day (intrinsic motivation). Thus, a product allowing a worker to work more efficiently when they want to (i.e. so they can finish early) may enhance the user’s experience.

It is not always the positive affective experience that is desired – at times, it is beneficial to experience negative affect for a net positive gain in the longer term (Jordan, 2002). Consider the Carmen Sandiego series of games. If they were too easy and did not provide challenge to the user, the game would probably not be as enjoyable because users would not feel the same sense of accomplishment when they progress. In the context of Carmen Sandiego, causing some user frustration and stress increased the overall positive experience, and subsequently the QIU was improved.

The idea that a negative affective response can have positive outcomes seems contradictory to hedonic theory, where enjoyment was the sum of the positive and negative experiences. By having these negative experiences, enjoyment should theoretically decrease. However, it must be realised that a minor negative experience may greatly enhance the positive experience that a user will ultimately have, and is therefore consistent with hedonic theory.

Johnson (1999) used computer games as an example of a product that emphasises fun. Using the Interacting Cognitive Subsystem (ICS) framework (Barnard and Teasdale, 1991), Johnson modelled user interactions with games and was able to provide insight into design aspects that made them fun. The key results were:

**List 3.2: Johnson’s Design Guidelines for Fun**

- Interface elements can convey affect and emotion in addition to information. Therefore, despite a game being technically inert emotionally, it is possible to simulate affect and emotion through the interface. This is the basis of all other results below.
- Associations between the interface and specific affects and emotions are developed through interaction. Using these elements consistently within the same game, and ideally throughout other games, will increase this association.
- Different users will have different associations depending upon their past experiences. Therefore, it is important to introduce these associations early.
and regularly in any game. Greater exposure to an association between an interface element and a specific emotion will be made stronger.

- Proprioceptive feedback (such as force-feedback controls) can increase a user’s arousal levels, but only if used consistently. Higher arousal levels can lead to heightened emotional experiences.
- Proprioceptive feedback can be associated with specific affects or emotions. As such, proprioceptive feedback is another interface element itself.

The associations discussed by Johnson are also the product of a “myriad of external, contextual factors including heating, lighting and noise levels” (Johnson, 1999). So while the results are beneficial to developers who wish to make associations between interface elements and specific affective experiences, they must still be viewed within the context of the overall system.

One element that has been used consistently on most modern computers is the ‘shut down’ or ‘off’ button (Figure 3.4). The affective experiences associated with the use of this button change depending upon the context in which it is being used. If the user is finishing their day at work, the button may be associated with the affective experience of relief, however if the user needed to use the button due to a product malfunction it may be associated with frustration.

![Figure 3.4: Stereotypical Shut Down or Power Off Button on most Computers](image)

There is increasing knowledge about product design and characteristics and their relationship to specific affective experiences. Both positive and negative affective experiences can have an overall positive influence on QIU.

It is, however, not possible to reliably relate specific design elements to a specific affective experience without accounting for the overall context of use. As Johnson’s (1999) guidance suggests the association between an interface element and an affective experience depends upon many factors beyond the scope of the design itself. However, it may be possible to identify common product characteristics that are typically associated with positive experiences (an idea explored in Chapter 6).
c. Interface Design Characteristics and Affect

“It is hard to find a good instance of harmonisation of beauty and rationale in the field of software engineering”, at least according to Matsubara (as quoted in Fujino, 1999). Most other development fields, such as architecture, have realised that systems have to do more than just function. Architects will first imagine the form of a building – the shape, the style – and will “choose the finest of their imagined concepts” and then design the required building to fit their vision of form in an effort to create something of beauty (Fujino, 1999).

To aid in the harmonisation of beauty and rationale within software of guidance has been provided within software engineering literature. The guidance includes descriptions on appropriate and appealing use of colours, fonts, screen layout, and other presentation elements (e.g. Constantine and Lockwood, section 3, 1999). In short, guidance has been provided to assist in making software aesthetically pleasing and ‘beautiful’.

According to Norman (1988), aesthetics, in the context of human-computer interaction, is often used as a marketing ploy – to get customers to like or want something even if it does not fit their needs. Aesthetics can ‘hide’ faults in a product, and when a product is aesthetically pleasing it is often at the expense of ease of use (Norman, 1988). Tractinsky (1997) has suggested that the neglect of aesthetics within systems development was to ensure that the focus of user-centred design remained on the user and their functional needs, rather than focussing on making product look good first and then worrying about the needs of users. While the necessity to address the needs of the user is important, aesthetics should be considered – and may even be a non-functional need of the user. Norman himself has pointed this out in rebuke of the many comments indicating that he believed in performance and productivity to the exclusion of everything else, including beauty, aesthetics, and other affective considerations:
My intention was not to denounce beauty. I simply wanted to position usability in its proper place in the design world: equal to beauty, equal to function: equal, but not superior. I neglected the topic of aesthetics because I thought it already well covered elsewhere. Unfortunately, my neglect was interpreted by many to imply that I was against beauty, for usability at all costs.

Users want and need more than a functional product. Users want and always have an affective and emotional connection with a product (Norman, 2004). The result of this connection will reflect upon the product negatively or positively. This is true in the field of architecture – people want to live and work in appealing atmospheres, not simply a functional atmosphere. This is also true in systems development. With the vast array of products to choose from it is often found that products, on a whole, will be similar in function and it will be ‘the little things’ or affective connections that will be the basis a final decision. For example, most mobile phones currently available are equal in terms of performance and productivity, yet users still have preferences (Lindgaard, 2001). Lindgaard asserts that these preferences are based, in part, upon the affective and emotional experiences users attribute to product use.

In an effort to understand the relationship between aesthetic appeal and usability, Tractinsky et al. (2000) varied the usability and aesthetic nature of two simulated ATM machines. They found that aesthetics had a greater influence on the perceptions of the product than did usability. When users were presented with an interface that was high in aesthetic appeal, users perceived it as highly usable regardless of actual usability levels, and participants tended to rate the product as appealing. Variations in actual usability levels had no effect on appeal of the interface or perceived usability. This led to the conclusion that judgements of usability and product appeal are predominately driven by aesthetics.

In a similar study, Lindgaard and Dudek (2002) varied the usability of a product, but kept the aesthetic appeal constant. In contrast to Tractinsky et al., they found that usability was the primary contributor to overall opinion of the product. Lindgaard and Dudek (2002) attributed the difference in findings between the studies to the severity of
usability problems encountered. In their study, users were unable to complete all tasks, whereas in the Tractinsky et al. study users were able to complete all tasks.

Both studies agree that usability and aesthetic appeal are independent constructs. However, these studies are not clear about respective roles that aesthetic appeal or usability have in the formation of user perceptions about a product. Combined, the results from the studies seem to indicate that both usability and aesthetic appeal can influence the user’s perceptions of a product, and therefore QIU. It should be noted that the studies by Tractinsky et al. (2000) and Lindgaard and Dudek (2002) both used performance-based tasks. It is unknown if the results found in these studies will apply to non-performance-based technologies and the QIU of these technologies.

This discussion on interface design characteristics and affect has been limited to aesthetics. Karvonen (2000) observes that research into the aesthetics of interactive products tends to be limited to the visual design of a product. However, in the foreseeable future, haptics, voice products, and other interaction methods will become more prevalent in design and there will need to be a more broad understanding of interface characteristics that contribute to affect beyond visual aesthetics.

d. Affective HCI Research Summary

Affective HCI looks at how the design of a product can create an affective experience in a user. While there has been little published research that targets the broad concept of affective HCI, there is a growing amount of literature within the sub-divided research areas (see Figure 3.3). Specifically, there has been:

- Exploration of psychological theories of affect and emotion and how these can apply to design,
- Research targeting specific user affects and how these can be elicited by a product,
- Investigation of visual aesthetics (as one product design characteristic) and the effects of aesthetics on user-experience and affect.

This review of affective HCI research has highlighted the actual breadth of the area. It seems that research into affective HCI is creating more questions than it is answering.
What started as a seemingly simple idea of “understanding how product design and interaction influence a user’s affect and emotion” (see introduction to section 3.3), is now a seemingly amorphous topic. However, it is the nature of science and research to ask more questions than those answered – particularly when an area of research is in its infancy like affective HCI. The affective HCI research that has been performed does highlight the potential influences of design on user affect, which in turn can influence the QIU experienced – particularly non-performance-based technologies.

3.4. Affect and Perceptions of Product Quality

QIU is the user’s perception of product quality, and is influenced by the internal and external qualities of the product, and how these are experienced by the user within a specified context of use (Bevan, 2000). In the context of performance-based technologies, QIU is concerned with the capability of a product to enable specified users to achieve specified goals with effectiveness, productivity, safety, and satisfaction in specified contexts of use (ISO/IEC 9126-1, 1991 as cited in Bevan, 2001).

Studies performed by Lindgaard and Dudek (2003) and Tractinsky, Katz, and Ikar (2000) each identified that the user’s perception of a product was dependent upon both the usability (related to performance qualities) and aesthetic appeal (related to affective qualities). Both studies required participants to perform tasks that were related to productivity and performance, and as such, it can be said that some performance-based technologies will, to some degree, be influenced by affective qualities. However, for non-performance-based technologies affect will have a more pronounced influence on QIU, as the affective experiences of the user are the major source of overall experience.

Kemp and van Gelderen (1996) explain that a user’s motivation to purchase and use a particular product is based on the subjective evaluation of both the utilitarian and experiential aspects related to the product. Depending upon the purpose of the product and the context of use, the weighting of these two aspects will vary. Some products – such as a screwdriver – will be based solely upon the utilitarian aspects of the product. Others – such as kitchen appliances – will have a combination of both the utilitarian and experiential factors contributing to the motivation to purchase the product. In addition,
some products – such as board games – place primary emphasis on the experiential qualities of a product.

Figure 3.5 relates the work of Kemp and van Gelderen (1996) to QIU. This model emphasises that QIU can be influenced by performance/productivity and emotional/affective aspects of a product. Each aspect is evaluated independently by a user, and the combined evaluation of these represents the QIU perceived.

![Figure 3.5: Quality in Use Duality in Performance and Experiential](image)

This diagram is not meant to imply equal weighting for the performance/productivity and emotional/affective aspects for all products. The weighting of these will represent the importance of these aspects upon a user’s personal evaluation of QIU, and this will depend upon the user’s needs and desires for a particular context. If the user wants to experience the affect of joy (for instance), then weighting for the emotional/affective side will increase, and vice versa. It is anticipated that for most products both performance/productivity and emotional/affective aspects will have some weighting associated with them. However, it is feasible to consider that the QIU experience for some products will relate only to a single aspect (i.e. QIU is based entirely on performance/productivity OR emotional/affective evaluation). For example, the user needs associated with a safety-critical system are likely wholly dependent on performance/productivity factors.

### 3.5. Chapter Summary

A user’s perception of a product can be influenced by the affective experiences during interaction with a product. Therefore, affect can influence QIU. With that said, little is known about the manner in which affect influences QIU.
This chapter introduced affect and how it relates to HCI and QIU. Specifically, this chapter:

- Offered detailed descriptions of affect and emotion, including operational definitions that can be used HCI.
- Distinguished affect and emotion, and highlighted how these distinctions can benefit HCI research and practice.
- Justified affect and QIU as the prime focus within this thesis (as opposed to emotion and QIU). The basis for this was that affect has a greater ability to influence product development by identifying specific elements of product interaction that contribute to negative or positive affective experiences.
- Summarised key research in the area of affect and design. This included summaries of research into affective computing (how can products recognise and react to user affect), and affective HCI (how to design to elicit user affect and emotion).
- Discussed how affect can influence QIU, particularly for non-performance-based technologies.

The three subsequent chapters outline studies that aim to further clarify the relationship between affect, design, and system quality. Chapter 4 explores satisfaction – a current element within QIU – as there have been suggestions that satisfaction and affect are the same. Chapter 5 continues by looking at the user-experience (which is sometimes equated to satisfaction and/or affect). Within this chapter a user-experience model is offered integrating user affect, user satisfaction, and user perceptions of product performance as they relate to different types of products. Chapter 6 reports on a survey, the results of which provide guidance to designers wishing to create positive affective experiences in users.
Chapter 4  Understanding Satisfaction

The previous chapter defined affect and discussed the potential benefits to product quality that can be gained by addressing affect during design. Further, it provided a review of research that has been ongoing in the area of affect and computers, with two main themes identified. The first theme was affective computing, which aims to create a machine that can identify and react to human affective states (see section 3.3.1). The second theme, and more pertinent to this thesis, is affective HCI research which focuses on the affective experiences that are elicited during interaction with a product (see section 3.3.2). This research can be considered in three parts (see Figure 3.3):

- Theoretical perspectives of design and affect
- Affective reactions and their relationship to product interaction
- Interface elements as they relate to specific affective responses

Affective HCI research is progressing and is increasing the understanding of the relationship between affect, design, and product quality. However, progress and acceptance of this research has been hindered by the ambiguous use of the terms affect and emotion. For this reason, affect and emotion were clearly distinguished, and the benefits of these distinctions for use within the context of HCI were summarised.

In section 3.3 it was suggested that QIU is derived from the subjective evaluation of both performance/productivity and affective/emotional experiences. QIU – as discussed in section 2.4.1.d – definitely addresses performance and productivity but does not have a clear link to user affective or emotional needs of the user. Some researchers suggest that the affect and emotion are addressed within the QIU attribute of user satisfaction, and that user satisfaction is representative of the overall user-experience (e.g. Grice, 2000). On the other hand, some researchers feel that satisfaction is used to describe any aspect of interaction not related to performance and productivity, and does not include
affect and emotion (e.g. Edwardson, 1998) – which effectively means affect and emotion would not be addressed as part of QIU measures.

The aim of this chapter is to clarify both the understanding of “satisfaction” and how this term relates to affect.

4.1. Satisfaction in HCI

Perhaps the most commonly used definition of satisfaction within HCI is “The comfort and acceptability of the work system to its users and other people affected by its use” (ISO 9241-11, 1998). However, within HCI related literature satisfaction has been used to describe concepts well beyond this definition. For example, Lindgaard (2001) highlights several unique uses of satisfaction within HCI literature, including:

- fun,
- enjoyment,
- aesthetic appeal,
- pleasure,
- user acceptance,
- comfort,
- user-experience, and
- adaptability.

With the variety of interpretations of satisfaction the boundaries have become blurred – satisfaction seems to be used as a ‘catch all’ term for anything that does not relate to another element of usability, system quality, or QIU. Indeed, Edwardson (1998) points out that feelings, emotions, affective experiences, and general likeability of a product have all been described under the guise of satisfaction.

These extended uses of “satisfaction” may be due to the increasing numbers of products that are addressing affective requirements, such as non-performance-based technologies. As discussed in section 2.7, non-performance-based technologies are not well addressed within the context of QIU, which is oriented towards performance-based factors. It may be possible that as researchers and practitioners came to realise that non-performance-based qualities are becoming important they adopted the only quality element that may
have some predictive and evaluative power for affect – satisfaction. However, using the term satisfaction in relation to fun, enjoyment, etc. adds ambiguity, and people will not immediately know what a specific person is referring to if he or she simply says “satisfaction”. To help illustrate the implications of the ambiguous use of “satisfaction”, consider this scenario:

*Two different word processors (Word-Type 1 and PrintThoseWords) have just undergone independent usability evaluations. The studies indicate that performance and productivity scores were comparable when objectively measured for each product. However, Word-Type 1 was found to provide users with high levels of satisfaction, and PrintThoseWords had marginal levels of user satisfaction. Which is ‘better’?*

It is not possible to conclude that Word-Type 1 is the better product, as it depends how the evaluators interpreted the word satisfaction. If the evaluators of Word-Type 1 interpreted it as user-acceptance and the evaluators of PrintThoseWords as fun, then it is impossible to compare results and know which word processor is better. Performance and productivity measures do not suffer this same issue because criteria and measures are well established and understood.

Allen and Buie (2002) recognised the issue of terminology being used in ambiguous and inconsistent manners within usability. As they eloquently state:

*We want a term to hold its value. It has to say the same thing to everyone who needs to read or hear it.*

and

*Semantics is all about the very meanings of the words we use, the intention in which we use them, and the understanding they create in our audience.*

If a word is used inconsistently or in an ambiguous manner then people may infer different meanings from what is said, or they may not be confident of their interpretation. Inconsistent or ambiguous use may also cast doubt on the validity and
reliability of research, as occurred in the 1940’s in psychology research related to affect and emotion (Tomkin, 1981).

In their article, Allen and Buie identified eight terms that are used inconsistently by HCI researchers and professionals. These were: intuitive; user-friendly; logical, heuristics; subjects; subjective; tester; and testing. Allen and Buie provided a clear and concise definition for each of these. They argued that in an area where practitioners deal with clients on a daily basis (who are usually not part of the HCI community) that the consistent use of terminology will increase the integrity of the field in the client’s eyes, and more broadly the wider population’s.

While Allen and Buie did not address “satisfaction”, their sentiments about the necessity of consistent use and meaning does have bearing on the use of the word. To ensure that we can communicate effectively – with other researchers, clients, or the public – “satisfaction” must be defined and used consistently within the context of HCI. The definitions offered by Allen and Buie were based upon their understandings of the term, and though they did justify their definitions there was no comparison with other available definitions. Unfortunately, “satisfaction” has been used in many ways and in many contexts, resulting in numerous understandings and definitions contributing to the ambiguity of the word. As such, to clearly define “satisfaction” a rigorous linguistic study into the etymology and lexography (meaning) of the term “satisfaction” was conducted. Etymology and how it applies to HCI is introduced in section 4.2, and the linguistic study is described in section 4.3.

4.2. Etymology, HCI, and Satisfaction

Etymology is the study of lexical history (Crystal, 2003). Etymology investigates the origins of specific meanings, how these meanings relate to each other, and how the word has come to achieve its present day meaning (Crystal, 2003). Etymology has traditionally focused on the study of word origins and how a word has evolved to its present state. Contemporary etymology also looks at groups of words that relate to a specific meaning.

Language and words are not static – these will evolve as a culture requires. Like a product, a language must meet the needs of the users, and as needs change language
must also change. New words are added (e.g. e-mail), antiquated words disappear from use, and some words will change meaning (e.g. the meaning of gay in the 1960’s compared to now). Semantic change – the process by which a word changes meaning – is considered part of the etymology of a word. Linguists identify four ways in which semantic change can occur (Crystal, 2003):

- **Extension or generalisation:** the meaning of a word broadens
- **Narrowing or specialisation:** the meaning of a word becomes more focused
- **Amelioration:** a word takes on a positive connotation (e.g. “lean” no longer refers to emancipation, but tends to refer to athleticism)
- **Pejoration or deterioration:** a word takes on a negative connotation (e.g. the origins of “villain” refers to peasants, not criminals)

Relating the word “satisfaction” to these four methods of change, it could be said satisfaction has been extended to have a larger range of meanings within the context of HCI. However, science aims to be precise and unambiguous and, to this end, meanings of words generally become narrower or more specialised (Crystal, 1997), and not generalised. If science were to accept the generalisation of words, this would be accompanied by a loss of precision, an increase in ambiguity, and deterioration in the ability to communicate and share knowledge succinctly.

There is a gap between everyday and scientific language (Crystal, 1997). The HCI community has strived to overcome this gap, and has made words and meanings used within HCI relatively consistent with the common present-day meanings known to the wider population. Where this has not been possible, clear definitions have been offered to the wider population so they know what a term means (e.g. Allen and Buie, 2002). In relation to satisfaction and HCI, there are two options:

- Use the meaning(s) of “satisfaction” as used in present-day English; or
- Provide the scientific community and wider population with a single clear definition of the meaning(s) of “satisfaction” within the HCI context.

Ideally, the scientific meaning of satisfaction within the HCI context should match that of present-day meaning commonly attributed to satisfaction. This will minimise the gap between everyday and scientific language, and will allow HCI practitioners to easily
relate to clients. However, within scientific disciplines it is not always possible to use the everyday definition of a word, as science requires a degree of precision that is not afforded in everyday language (Crystal, 1997).

A third option is to accept fun, enjoyment, comfort, acceptability, adaptability, user-experience, etc. all as unique meanings of satisfaction. However, this option should not be considered, as it would be impossible to understand what a HCI practitioner or researcher means when they use the term satisfaction. Further, accepting all meanings as part of satisfaction may serve to alienate the wider population, with whom the HCI community deal with on a regular basis. That is, if a practitioner tells a client that they evaluate the experience of fun and then proceed to evaluate user satisfaction the client may become confused because he or she may not relate fun to satisfaction.

By way of contrast, the acceptance and adoption of these many interpretations of satisfaction may actually reflect a change in everyday use and meaning of the term satisfaction. That is to say, perhaps the present-day use of “satisfaction” does include fun, enjoyment, etc. If this is true, then the acceptance of these within the scope of satisfaction in HCI may be justified.

The following study aims to determine what the current everyday meaning of “satisfaction” through the use of a lexographic tool – WordNet.

4.3. Study: What is Satisfaction?
   A condensed version of this study was published in Bentley, Johnston, and von Baggo, 2002c

The varied and inconsistent use of the term satisfaction has led some people to ask the question what satisfaction actually means within the context of HCI. For example, Karvonen (2000) asked, “What is meant by satisfaction?”, and Lindgaard and Dudek (2003) asked – and explored in detail – “What is this evasive beast we call user satisfaction?”

Lindgaard and Dudek (2003) reported on three experiments that aimed to increase the understanding of the relationship between satisfaction and the user-experience. In each experiment performed, participants used a website, provided subjective commentary throughout use, and completed the Website Analysis MeasureMent Inventory
(WAMMI, 2004). The aesthetics and usability of these websites varied within each experiment. Based on the combined results of these studies, Lindgaard and Dudek concluded that:

\{satisfaction\} is a complex construct comprising several affective components as well as a concern for usability, and that a priori expectation seem to play a major role in shaping user satisfaction. The concepts of aesthetics, emotion, expectation, likeability, and usability, all appear to influence the interactive experience, but they vary in prominence according to type of experience.

This description suggests that the varied interpretations of satisfaction – including fun, enjoyment, comfort, appeal, etc. – can influence satisfaction, but are not satisfaction themselves.

### 4.3.1. Purpose of Study

While the description of satisfaction offered by Lindgaard and Dudek (2003) highlights the many influences on satisfaction, it does not actually describe what satisfaction is or what it means to be ‘satisfied’ with a product. The study described here researches the etymological basis of the word satisfaction within the English language. The aim of this study is to answer the following questions:

- What is the present-day commonly understood meaning of “satisfaction”?
- How does the present-day meaning of the word satisfaction compare to the varied meanings and interpretations within HCI literature?

By asking these questions, it will be possible to understand whether the HCI community has generalised the use of satisfaction beyond what is commonly used in present-day English. This study will also provide insight into how satisfaction relates to affect.

### 4.3.2. Method

This study explores the etymology of ‘satisfaction’ in two ways:

- Etymology (origins and evolution of the word satisfaction)
• Lexography (looking at the contemporary meanings of the word satisfaction, and how they are related to other words or terms)

The two approaches allows for the validation of results. That is, if both approaches provide the same results, then they can be considered to validate each other.

To study the etymology of satisfaction, the Complete Oxford English Dictionary (Simpson and Weiner, 1989) was consulted. In addition to word meanings, this dictionary provides a history of words including the literal word origins and the date of first appearance of a word. It also provides a date of appearance for new or changed meanings. Results from this review is provided in 4.3.3.

To study the lexography of satisfaction, a tool called ‘WordNet’ was used. WordNet is a computer-based lexical database of the English language and allows the exploration of relationships between word meanings and other words or terms (Miller, 1991; see section 4.3.2.a for a detailed description). This provides an understanding of the contexts in which the different meanings of a word are used. The results from this analysis are provided in section 4.3.4. WordNet is based on the Brown corpus, which is a ‘carefully selected body of literature’ which represents present-day English use (Francis and Kucera, 1979).

The results from the two analyses are then compared for consistency and integrated (section 4.3.5). Given that the two study methods are linguistically based, the results should be similar and can be used for cross-validation.

A discussion follows in section 4.3.6, and an interpretation of satisfaction for use within the context of HCI that acknowledges the distinction between affect and satisfaction is provided in 4.4.

a. Tools — WordNet (a Lexical Database)

A dictionary provides a mapping of words to their meanings, and a thesaurus identifies words that have similar meanings. However, the alphabetic structure of dictionaries and thesauri do not reveal anything about the contextual meaning of words (Miller, 1991). Alphabetization is a convenient and practical way of organising words but results in the scattering of words in a non-contextual manner (Miller, 1991). That is to say, two
adjacent words in a dictionary do not necessarily have any relationship to each other, except for being alphabetically close.

A lexical database provides a cluster of concepts related to a word and its meaning. A lexical database is compiled by expert lexographers, who interpret and identify the contexts, the meanings, and uses of words in everyday language. They use a ‘corpus’, which is effectively an extremely large sample of recent text (including transcripts of speech) which is said to represent modern day language (Miller, 1991). Lexical databases can be used as dictionaries or thesauri, but the real benefit of a lexical database is the ability to relate words by context of use and meaning. A lexical database will allow one to relate words at various levels, making the context of use apparent. For example, a lexical database will provide details of hypernyms and hyponyms for every sense of a word. In the simplest terms, the sense of a word can be considered a unique meaning of that word. Hypernyms and hyponyms are:

- **Hypernyms** are higher order words that provide a more generic concept than the word being investigated (i.e. <word> is a kind of …; a parrot is a kind of bird).
- **Hyponyms** relate to lower order words, which place additional details in the meaning and are more restrictive than the original word (i.e. <word> contains the concepts of …; run contains the concepts of marathon, track event, sprint, etc.).

One lexical database in widespread use is WordNet (Miller, 2002). WordNet is used by several dictionaries/thesauri (e.g. Webster’s, Roget’s) as the source of their definitions. To provide some sense of context dictionaries and thesauri often provide sample sentences, however these sentences convey only a minimal amount of contextual meaning compared to a lexical database.

**b. Procedure (for WordNet Analysis)**

To begin, the term ‘satisfaction’ was entered into WordNet (PC version) to discern what the word means in the English language. Variations of the word satisfaction were also entered (e.g. satisfy) to ensure that the analysis incorporated all possible interpretations. This step provided the unique senses (meanings) of satisfaction.
After the unique senses of satisfaction were determined, hypernyms and hyponyms for each sense of satisfaction were extracted from the lexical database. These provide a context to each sense, allowing for a greater understanding how the term satisfaction is conventionally used in the English language. The hypernyms and hyponyms were used to build a contextual map for the unique senses of satisfaction.

4.3.3. Historical Etymology Results

The historical etymology was obtained through review of the Complete Oxford English Dictionary. According to that dictionary, the word satisfaction is a combination of the Latin terms *satis* (sufficient) and *facere* (to make), with a literal translation of “to make sufficient”. A historical timeline was developed highlighting key dates when unique uses or different contexts of satisfaction emerged (shown in Figure 4.1). The figure shows the emergence of satisfaction uses to the nearest century (the Complete Oxford English Dictionary provides exact years of emergence). The distinct meanings were formed into five related groups of meanings. These referred to:

- **Group 1**: The fulfilment of obligations or duties (i.e. payment in satisfaction of a debt)
- **Group 2**: Dignity and honour (i.e. to demand satisfaction in a duel)
- **Group 3**: Desires or general feelings (i.e. to fulfil one's hunger is to satisfy one's hunger; to have the satisfaction of meeting someone)
- **Group 4**: Information removing doubt or fulfilling a demand or need (i.e. to be satisfied with the decision in a court case)
- **Group 5**: Acknowledgement of a debt being paid (i.e. to enter into satisfaction that a debt has been paid)

According to the Complete Oxford English Dictionary, each meaning of satisfaction is still found in the English language today with the exception of the final meaning.
Figure 4.1: Historical Etymology Timeline of Satisfaction
4.3.4. Meaning (WordNet) Etymology Results

The WordNet analysis revealed four distinct senses (meanings/contexts) of the word satisfaction. These four senses were (these senses are numbered according to their order of appearance from WordNet):

- **Sense 1**: satisfaction (the contentment you feel when you have done something right; "the chef tasted the sauce with great satisfaction"")
- **Sense 2**: gratification, satisfaction -- (state of being gratified; great satisfaction: "dull repetitious work gives no gratification")
- **Sense 3**: atonement, expiation, satisfaction (compensation for a wrong; "we were unable to get satisfaction from the local store")
- **Sense 4**: satisfaction, satisfying (act of fulfilling a desire or need or appetite; "the satisfaction of their demand for better services")

Hypernyms and hyponyms were extracted for each sense, allowing a contextual map of the word satisfaction to be created, as shown in Figure 4.2. Hypernyms are shown above the sense, and hyponyms below. Words in **bold** represent ‘unique beginners’. Unique beginners are very important in the development of a lexical database as they represent the starting point upon which all other words and contexts are based. However, within the context of this study unique beginners do not have any additional meaning or effect on results aside from being a word or context with no hypernyms.
4.3.5. **Comparison and Integration of Results**

The results of the two analyses revealed similar meanings and uses of the word satisfaction in history and in everyday language. A complete comparison is shown in Figure 4.3.
<table>
<thead>
<tr>
<th>Group 1:</th>
<th>The fulfilment of obligations or duties (i.e. payment in satisfaction of a debt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense 1:</td>
<td>satisfaction (the contentment you feel when you have done something right; “the chef tasted the sauce with great satisfaction”)</td>
</tr>
<tr>
<td>Group 2:</td>
<td>Dignity and honour (i.e. to demand satisfaction in a duel)</td>
</tr>
<tr>
<td>Sense 1:</td>
<td>satisfaction (the contentment you feel when you have done something right; “the chef tasted the sauce with great satisfaction”)</td>
</tr>
<tr>
<td>Group 3:</td>
<td>Desires or general feelings (i.e. to fulfil one’s hunger is to satisfy one’s hunger; to have the satisfaction of meeting someone)</td>
</tr>
<tr>
<td>Sense 3:</td>
<td>atonement, expiation, satisfaction (compensation for a wrong; “we were unable to get satisfaction from the local store”)</td>
</tr>
<tr>
<td>Group 4:</td>
<td>Information removing doubt or fulfilling a demand or need (i.e. to be satisfied with the decision in a court case)</td>
</tr>
<tr>
<td>Sense 2:</td>
<td>gratification, satisfaction -- (state of being gratified; great satisfaction: “dull repetitious work gives no gratification”)</td>
</tr>
<tr>
<td>Group 5:</td>
<td>Acknowledgement of a debt being paid (i.e. to enter into satisfaction that a debt has been paid)</td>
</tr>
<tr>
<td>Sense 4:</td>
<td>satisfaction, satisfying (act of fulfilling a desire or need or appetite; “the satisfaction of their demand for better services”)</td>
</tr>
</tbody>
</table>

Figure 4.3: Comparison of Grouped Meanings and Senses of Satisfaction

Note that the senses are not in numerical order. This is done to simplify the figure.

Group 5 provides the only inconsistency as it is not related to any sense of the word satisfaction as extracted from WordNet. As previously mentioned, group 5 represents a specialised use of the term satisfaction in a legal context that is no longer used according to the Complete Oxford English Dictionary, and as WordNet is based on the everyday use of language this use does not exist in the database. Excepting this, the groups and senses are consistent suggesting that the results from the analyses are reliable representations of how the word satisfaction is used in everyday English language. That is, the historical progression and development of the word satisfaction (as obtained through the etymology) is consistent with the current uses.

The most significant outcome of this study was that the word satisfaction always relates to the achievement or fulfilment of goals or obligations. Senses 1 and 2 relate to the general feelings that an individual has when completing a goal or after the goal has been fulfilled. Senses 3 and 4 relate to the actual act of satisfying and fulfilling goals (rather than the feelings as in senses 1 and 2).

4.3.6. Discussion of Results

Allen and Buie (2002) stress the importance of consistent use of terminology, particularly in a field such as HCI with people from a wide variety of backgrounds. If a
word is used inconsistently, it may diminish the credibility of the person using it and the overall field that they represent.

While it was considered possible that the everyday meaning of satisfaction has expanded to include the varied interpretations cited by Lindgaard (2001), the study revealed that this is not the case, and that “satisfaction” is being used inconsistently. The expansion of meanings attributed to satisfaction seems to have been driven by the HCI scientific and research communities. Expanding a definition within science is generally considered counter-productive, as it decreases the ability to communicate and share knowledge effectively. Within a science terms and definitions become narrower as the knowledge about a specific concept increases (Crystal, 1997), and this enhances the degree of precision and clarity of communication and knowledge sharing.

The results of this study suggest that these interpretations of satisfaction are exceeding the boundaries of its meaning. This supports Lindgaard and Dudek’s (2003) conclusion that “satisfaction” is being used in HCI in a manner inconsistent with its meaning. Further, the varied interpretations may actually be hindering scientific progression, as it is not always clear what researchers mean by “satisfaction”, and therefore it is not always possible to build upon existing research effectively.

The fact that the HCI community deals with the wider population on a regular basis means that the language used within HCI should attempt to be consistent with everyday meanings. The study revealed that “satisfaction”, in everyday use, deals with the general feeling related to goal accomplishment or to the act of completing a goal (i.e. the act of satisfying a need). Using interpretations that are not consistent with this everyday use may confuse the wider population, and may make them wonder about the credibility of the HCI field. It is doubtful that the interpretations attributed to “satisfaction” within the HCI literature will become part of everyday English. Therefore, the HCI community should accept and use “satisfaction” in a manner consistent with everyday use, and the concepts that are not addressed within this should be labelled differently.
4.4. An Interpretation of Satisfaction for HCI

The etymological and lexographic study revealed that satisfaction has two very broad interpretations. These are:

- The general feelings one has when either striving to achieve or after achievement of a goal (e.g. getting satisfaction from ones work; to be satisfied with a job well done)
- The act itself of achieving or fulfilling a goal (e.g. to satisfy a demand, hunger, or need)

The first interpretation is similar to the existing definitions of satisfaction within the HCI context. For example, the definition of satisfaction offered by ISO 9241-11, “The comfort and acceptability of the work system to its users and other people affected by its use”, is related to the feelings workers have towards a system. The second interpretation related the act of fulfilling goals is addressed within the aspects of performance and productivity.

As Lindgaard and Dudek (2003) concluded, satisfaction can be influenced by performance-based and non-performance-based factors, and this conclusion is confirmed by the two interpretations of satisfaction derived from this study. Depending on the type of product being used, the degree of these influences will vary. Based on the findings of Lindgaard and Dudek (2003) and the results of this study, the following definition of satisfaction is proposed for HCI contexts:

**Definition 4.1: Satisfaction**

*Satisfaction is the general feeling experienced when striving to achieve or subsequent to the achievement of goals that are considered important to an individual. These goals may be externally imposed (e.g. work goals) or they may be related to personal, internally imposed goals.*

The inclusion of goals that are important to an individual is necessary. A person who is achieving goals that mean something – whether performance or non-performance related – can develop satisfaction. A person who is using a product without caring about the result is less likely to develop any sense of satisfaction.
This definition does not include affective experiences such as fun, enjoyment, or frustration, as the etymological study showed that these concepts are not part of the everyday meaning of satisfaction. However, the definition does not preclude the possibility that affective experiences such as these can influence satisfaction. If a person has an affective experience while interacting with a product (whether it is performance-based or non-performance-based) it may influence the degree of satisfaction that one has. It is possible that, for some products, satisfaction is the result of affective experiences. For example, if the user wishes to experience enjoyment while using a product, and they achieve this, then this may contribute to the individual’s satisfaction. In this respect, this interpretation is consistent with Lindgaard and Dudek’s (2003) description of factors influencing satisfaction. (For ease of reference, their description is copied below and was first introduced in section 4.3).

{satisfaction} is a complex construct comprising several affective components as well as a concern for usability, and that a priori expectation seem to play a major role in shaping user satisfaction.

Satisfaction cannot be considered an affective experience itself because, by definition, an affective experience is a short-term, discrete feeling, and is not subject to external influences. The duration of satisfaction can vary, and is subject to many influences as discussed by Lindgaard and Dudek (2003). Satisfaction could potentially be considered an emotional experience (as suggested by sense 2 in Figure 4.2).

The interpretation of satisfaction provided in Definition 4.1, combined with Lindgaard and Dudek’s (2003) description of factors that can influence satisfaction provides an inclusive understanding of satisfaction within the HCI context. Using the proposed definition of satisfaction will reduce and potentially eliminate scenarios such as the one above. Further, restricting the meaning of satisfaction within the context of HCI will increase the ability to communicate clearly, succinctly, and with the degree of precision that a science demands. This definition of satisfaction also means it is necessary to use different terms that relate to affective experiences, as these are not part of satisfaction (although they can influence satisfaction).

Overall, this definition “satisfaction” will benefit designers, users, researchers, the general HCI community, and the public who have dealings with the HCI community.
4.5. Chapter Summary

The word “satisfaction” has been used inconsistently within HCI literature. It seems as though “satisfaction” is used to describe any interaction property that does not directly relate to performance or productivity. For example, it has been used in reference to affective experiences, comfort, acceptability, user-experience, etc. However, these uses seem to extend beyond established definitions of satisfaction.

This chapter explored what “satisfaction” actually means in everyday language and within the context of HCI, and how these meanings relate to affect. Specifically, this chapter:

- Reviewed various uses and definitions of “satisfaction” within the HCI literature.
- Argued that satisfaction needs to be used consistently, particularly as HCI professionals deal with the wider population on a regular basis.
- Suggested that the ideal interpretation of satisfaction is one that is consistent with that of the present-day English use, and is precise enough to be useful within the context of science.
- Reported on a linguistic based study that revealed:
  - The present-day use and interpretation of the feeling of satisfaction is related to the achievement or process of achieving goals that are important to an individual.
  - The interpretations of satisfaction within HCI literature does extend beyond the everyday use, and is in fact more general than everyday use of the term (which effectively reduces the scientific value of the term satisfaction).
  - That affect and satisfaction are distinct concepts, although affect can influence feelings of satisfaction.
- Proposed an interpretation of satisfaction for use within the context of HCI, based upon the results of the linguistic study and findings by other researchers.
The study and discussion of satisfaction led to the conclusion that user satisfaction is a general feeling that is derived from achieving goals that an individual considers important. In their description of influences on satisfaction, Lindgaard and Dudek (2003) say “the concepts of aesthetics, emotion, expectation, likeability, and usability, all appear to influence the interactive experience, but they vary in prominence according to type of experience.” This raised the notion of the interactive- or user-experience. Lindgaard and Dudek (2003) felt that there was some relationship between the user-experience and satisfaction. However, they did not explore or suggest what this relationship was.

The user-experience appears to have many commonalities with QIU. Certainly, by the very nature of the term, the user-experience relates to the experience a user has when interacting with a product. If the user’s needs and/or desires are sated then the user will normally have a positive experience. QIU has been described as the capability of a product to allow a user to achieve specified goals (see 2.4.1.c for a more detailed description). For QIU, if a user can achieve their goals efficiently, effectively, etc. then the user will normally perceive high QIU. Because of these similarities, it was considered prudent to explore user-experience, and how knowledge and ideas related to the user-experience could help the integration of affect into the concept of QUI.

This chapter explores the knowledge of the user-experience, and how this relates to affect and QIU.

5.1. Perspectives of the User-Experience

“User-experience” has become a buzzword within business, manufacturing, and service provision. Each want the users of their products to have a positive experience – assuming that this means the user will buy the product, use the product again, or tell
their friends to buy the product. For example, Optus (an Australian telecommunications company) has recently prepared a business readiness trial for digital television rollout. In the introduction to the participants guide to the trial, it says (Optus, 2005):

\begin{quote}
The purpose of the business readiness trial is to ensure that the overall customer experience is positive prior to the launch of Optus TV featuring FOXTEL digital
\end{quote}

As a result, the user-experience has become a development aim within HCI. Unfortunately, there is debate regarding what the user-experience actually is. Four different perspectives of the user-experience are discussed below (holistic, performance, product-centric, and combined performance and non-performance).

\section*{5.1.1. A Holistic View of the User Experience}

Hesketh.com (2002) describes the user-experience as “a way to describe users’ successes, failures, and thoughts about these events as they browse or complete tasks”. In further description, Hesketh.com recognises that everything surrounding a product will impact the experience. This includes branding, attitudes towards the makers of the product, the context of use, etc.

This holistic view of the user-experience may be correct, but it does not provide any guidance to a developer wishing to create a positive user-experience. Telling a developer that ‘everything’ impacts the user-experience will not provide suggestions on how to create a positive experience. Further, evaluating the user-experience from this holistic perspective is impossible as there many variables to account for outside of the product itself.

\section*{5.1.2. A Performance View of the User Experience}

Another view of the user-experience relates solely to performance of a product. The basis for this view is if the product performs adequately then the user will have a positive experience. Nielsen (1993) emphasised this view when he advocated product performance above all else. However, Nielsen has since admitted that he asserted this too strongly, and that he does recognise that there is more to the user-experience than product performance (Nielsen, 1996). Nielsen still argues that usability and product
performance are the primary indicators of a positive user-experience, and that if these are lacking then the user-experience will suffer.

Davis, Bagozzi, and Warshaw (1992) provide support for this perspective. They conducted a study to determine what motivates a person to use a specific product within the context of a workplace. They found that the primary motivator was whether the product will aid the worker in the performance of their tasks. Other factors – such as aesthetics, fun – can only add to user-experience when the performance aspects are catered for (i.e. if two products performed equally within a given context, then users would choose the one that appeals to them the most). However, the study by Davis, Bagozzi, and Warshaw should be considered relevant only to the context of the workplace, and their results should not be generalised beyond this context.

While there is recognition that the user-experience is more than just product performance (e.g. Nielsen, 1996), performance measures are sometimes used as the only user-experience measures. For example, Schmitt (2001) states that the user-experience of a web site can be measured by reviewing logs of users (hits, return visits, number of sessions, number of unique visitors, length of sessions, number of pages viewed per hit, number of sales, etc.). This certainly measures the use of the website, and can provide details of the business experience (e.g. was a sale achieved?), but it does not address what the user actually experienced.

5.1.3. A Product-Centric View of the User Experience

Garrett (1993) suggested that good coding, sharp graphics, providing the required functions, and making the product aesthetically appealing will create a positive experience in the user. This view of the user-experience analogous to the idea of product quality – Garrett has essentially said that high internal and external quality will create a positive user experience. Internal quality would involve good coding, sharp graphics, etc. and external quality involves the functional behaviour of the product. However, this view is product-centric, and neglects the experience that a user has within their context of use. It should be noted that this view was written in 1993, and was typical of views of the system development process at that time.
5.1.4. A Combined View of the User Experience

Whitten (2001) sees a positive user-experience as being afforded by three main characteristics – usefulness, usability, and aesthetics (see Figure 5.1). He believes that if a product incorporates each of these well, then it will also create a positive user-experience. His model provides distinct elements that can be addressed or included in the design to meet each of the characteristics – for example reducing errors will increase usability.

Within this model, “fun” is contained within aesthetics. However, “fun” and other affective experiences are not necessarily derived from aesthetics. For example, early text based computer games have very poor aesthetics, but were (and still are) considered fun. Regardless, the model does support the inclusion of both performance- and non-performance-based factors within the concept of user-experience.

5.1.5. Summary of Views

These four perspectives on the user-experience outline the difficulty that developers have when asked to target the user experience. Should they target performance? Aesthetics? Or is there something else? The view offered by Hesketh.com is perhaps the ‘safest’ – it is impossible to be wrong when suggesting that anything could affect the
user-experience. However, this view does not help developers know what they should actually do.

The other three views offered do provide some guidance. They describe characteristics within a product or behaviours of a product that should create a positive user experience. Each has limitations however. The performance-oriented view neglects the many products where users just want to have fun (for example). The product-centric view discussed neglects non-functional needs of the user and the context of use. The combined view addresses the limitations of the other two, but is limited by not identifying that different products and different users will have different needs and desires for a product.

5.2. Proposed Model of the User-Experience

Both performance-based and non-performance-related factors can influence the user-experience (Nielsen, 1996; Whitten, 2001). The importance and weighting of these factors depends upon the product and the context of use. If the user requires a product to help them complete a task then the performance related factors may hold greater weighting in the development of the user-experience. Lindgaard and Dudek (2003) seem to agree with this as they discuss the variety of influences – some performance and others non-performance related – and indicate that their influence on the interactive experience depends on the type of experience sought.

The context is very important to the user-experience. The user’s experience depends on what they want from a system at the particular time. One method of differentiating context, as used throughout this thesis, is the relative performance orientation of a product – is it performance-based or non-performance-based? Figure 5.2 is a schematic of the user-experience that highlights the differing contributions from performance or affective related factors for different degrees of performance orientation. For workplace products performance and productivity factors are the prime determinant of quality, and as such these elements dominant the user-experience. Within the context of computer games it is the affective factors which dominate, and the performance of the product are of negligible importance.
Within this schematic, the affective factors address the concepts defined within the scope of affective HCI (discussed in section 3.3.2). This includes the actual affective experiences (e.g. fun), the design of a product that may enhance these experiences (e.g. aesthetics), and may include the application of related theories in order to enhance the affective experience.

Two important properties of this schematic should be noted. First, satisfaction is modelled between performance/productivity and affective factors as satisfaction can be influenced from either of these. Secondly, even for extremely performance-based products affective factors still contribute to the user-experience (though to a small degree) – and vice-versa for extremely affective-based products. This relates back to the discussion on prioritisation and trade-offs (section 2.4.2), where the onus is on the developer to determine the appropriate amount of resources to commit to affective or performance-based factors.

Though this purely a schematic view, the properties have been confirmed by Lindgaard and Dudek (2003) who have found that the user-experience is based on both performance- and non-performance-based factors. The schematic shown in Figure 5.2 has been kept simple and does not expand upon specific affective or performance related factors. This allows the schematic to remain representative of the user-
experience as we gain a greater understanding of affect and performance factors as they relate to the overall user-experience. Further, as the schematic is a representation of the user-experience across a range of products, it is impossible to highlight the relative contributions of individual factors (e.g. simply because aesthetics play a role for one product it does not necessarily mean it will play a role in another). The contributions of the specific individual factors towards the user-experience will change depending on the context of use. The schematic uses workplace products and computer games as examples of the two extremes of performance orientation. Most products will be somewhere between these two examples, with the relative contributions from performance and affective factors changing as appropriate.

The schematic should not be taken literally. In reality, few products would actually have a weighting of factors such that they would fall exactly on the graph. A more accurate representation of the user-experience could be given by the summative formula of:

\[
\text{User-experience} = \sum w_i \times m_i
\]

\(w_i\) is a weighting for the \(i^{th}\) factor
\(m_i\) is the metric for the \(i^{th}\) factor

This formula is a re-interpretation of the product quality formula offered by McCall et al. (1977) (recall Formula 2.1, page 16), changing the focus to the factors that influence the user experience rather than product quality, although some factors would contribute to both. Formula 5.1 allows for the key user-experience factors for a specific product and context to be included in the equation, along with appropriate weightings. These may involve a myriad of performance-, affect-, or satisfaction-related factors and related metrics. The exact choice of factors and metrics will depend on the product and context of use, and this formula provides for the inclusion of any possible influencing factor.

Formula 5.1 is a complimentary representation of the user-experience to Figure 5.2. Figure 5.2 is excellent at showing the general relationship between affective factors, performance factors, and satisfaction and their contribution towards the user-experience in different contexts of use. The strength of this figure is its ability to convey the key points about this relationship easily; however, the detriment to this is that it is
simplified. Formula 5.1 does not show a relationship between user-experience factors and context of use, but does allow for the precise contributions of each factor to be included. The combination of Figure 5.2 and Formula 5.1 provide a good overall representation of the user-experience.

5.3. User-Experience and Quality in Use

The user-experience can be considered the ‘HCI equivalent’ to the system quality concept of QIU. The definition of QIU (originally provided in section 2.4.1 and copied below for ease of reference) correlates well with the representation and understanding of user-experience discussed.

**Quality in Use:** the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety, and satisfaction in specified contexts of use


Both quality of use and the user-experience is a reflection of a product’s ability to fulfil the users’ needs or wants within a specific context of use – QIU could be considered capability of a product to instil the appropriate user-experience in a given context. The main issue with the understanding of QIU is the neglect of affect-related factors as effectiveness, productivity, and safety are all related to the performance aspects of a product (refer to section 2.7 and 2.8 for elaboration on this). The only attribute that is, in any way, related to affect is satisfaction. However, as described in section 4.4, affect can influence satisfaction but affect and satisfaction are distinct concepts.

The user-experience is, literally, the experience that the user has with a product in a specific context. Bevan (2000) has described QIU similarly, saying that QIU is the user’s perception of the quality of the product and how these qualities are experienced within a specific context. This re-affirms the point that QIU and the user-experience represent the same underlying concept.

As QIU and the user-experience can be considered equivalent, the schematic shown in Figure 5.2 can be considered representative of QIU. Further, Formula 5.1 will also apply to QIU. Re-writing this formula to include QIU produces:
Formula 5.2: User-experience and Quality in Use

\[ \text{User-experience} = \text{Quality in Use} = \sum w_i \times m_i \]

- \(w_i\) is a weighting for the \(i^{th}\) factor
- \(m_i\) is the metric for the \(i^{th}\) factor

Using the schematic shown in Figure 5.2 and Formula 5.2 will allow the designer of a product to make the judgement regarding the important factors contributing to QIU. This judgement should be based on knowledge about the product, the users, and anticipated contexts in which it will be used. The use of the formula also emphasises that each factor considered important to QIU should also be measured in some way.

5.4. Chapter Summary

Within the discussion of satisfaction (Chapter 4), the notion of user-experience appeared. User-experience appeared to have many commonalities with QIU, and it was considered prudent to explore how user-experience knowledge could benefit QIU.

In addressing the user-experience, this chapter:

- Reviewed various perspectives of the user-experience. These included holistic, performance, product-centric, and combined performance and non-performance perspectives.
- Suggested that the user-experience is a combination of performance factors, non-performance factors, and user satisfaction. A model was also offered emphasising that the relative contributions of these factors will depend upon the context within which the product is being used.
- Compared user-experience and QIU, finding that they each described the same underlying concept.

Thus far, this thesis has provided background to system quality, affect, and terms that have been incorrectly equated as affect (e.g. satisfaction, user-experience). The following chapter is more practical in nature for the developer. It reports on a survey relating three theories (and their individual elements) to positive user affect. These three theories have all been said to relate design and positive affect. Based on the survey results, guidance is offered to the developer who wishes to enhance the positive
affective experiences of users. Chapter 7 and Chapter 8 are also practical in nature, discussing methods that could be used to validate user affect as a component of QIU.
Chapter 6  *Exploring the Relationship between Affect and Design*

Designers, researchers, and other product stakeholders are seeking a greater understanding of affect and its relationship with design and system quality. Within the HCI literature, there are many new articles by leaders in the field, each lending their support to affect and how it may impact upon the design and success of a product.

Previous chapters have highlighted the impacts that affect can have on system quality, and have reviewed affect-related literature to increase the overall understanding of affect in relation to product design and system quality.

One concern regarding affect and its relationship with systems development is that affect is an ‘unknown quantity’ in design – that is, there is a lack of guidance available to designers to assist them in creating appropriate affective experiences in the user. Section 3.3.2 outlined some literature that related design and affect. However, much of this research was limited in scope focusing on specific elements of a design or on specific affects, rather than trying to understand the overall relationship between affect and design. The only research found that does address this overall relationship is the work by Johnson (1999) (see List 3.2 on page 58). Having an understanding of the overall relationship between affect and design will provide options and alternatives during the design process; giving the designers the opportunity to choose what they believe will best elicit the desired affective experiences during product use.

This chapter investigates the holistic relationship between affect and design. It begins by identifying some key issues in studying the relationship between affect and design – namely the need to choose an appropriate product to study. The chapter continues by providing background to and results of a web-based study that explores theories that have been said to relate affect and design. The aim of this was to establish what, if any, theory or aspects thereof provide a broader understanding of the affect-design
relationship. Based upon the study results, some guidance is offered to designers who wish to create an affective experience.

6.1. Choosing a System to Study Affect

It is rather obvious to state that in order to evaluate user affect it is necessary for the user to experience affect. Therefore, within the context of system use it is necessary to select a system that aims to and will create affective responses.

Ideally, affect should be studied in a system where affective experiences are considered important, such as non-performance-based technologies. Although people may, at times, interact with performance-based products on an affective and emotive level (e.g. getting frustrated), this is not a primary goal of the system or user – rather it is a side effect (sometimes undesirable) of product interaction. Another benefit to studying affect in non-performance-based technologies is that they can be expected to elicit a greater number of user affective experiences, and generally, these experiences will be stronger, making them easier to identify and measure. Affective experiences in performance-based systems may be overlooked because the degree of the experience may be insufficient to be observed and measured.

One non-performance-based technology that can be used to effectively study affect is that of computer games. The aim of a game is to create positive affect within a person, which then leads to an overall positive experience for that person (Csikszentmihaly, 1992). Computer games target user affect, and as Draper (1999) suggests computer game users rarely expect anything else other than the experience of playing. As such, computer games are the ‘model’ non-performance-based technology to study affect, as the quality is derived almost wholly from affect-related factors (recall Figure 5.2, page 89).

Computer games are also known to be extremely immersive, with users often saying that they forget about the external world while interacting with the game (Csikszentmihaly, 1992). As such, the affective reactions that users experience during interaction with a computer game are a product of the interaction, and not influenced by external variables that could be a factor when a user is not fully immersed in an activity. It can also be assumed that computer games will elicit frequent affective responses and
thus create more opportunities to observe affective responses in situ. Both the immersion and greater number of affective responses make computer games an ideal product to use when studying affect.

To summarise, to study affect in the context of system use it is important to choose a system where affect is considered important, such as non-performance-based technologies. Computer games provide the opportunity to study the relationship between affect and design with minimal impact by external variables. It is for this reason that many researchers have used computer games within their work – either as case examples (e.g. Draper, 1999), or as a studied product (e.g. Johnson, 1999). This is also the reason computer games were used to study several theories reported to relate affect to design (study is in section 6.3), and to explore methods to evaluate affect (section 8.3).

6.2. Theories Relating Affect and Design

While research is limited in regards to the holistic relationship between affect and design, there are existing theories that have been suggested, to some degree, as describing this relationship. Three of the most commonly suggested theories include:

- Flow (section 6.2.1) – as suggested by Draper (1999)
- Heuristics for internally motivating interfaces (section 6.2.2) – as suggested by Malone and Lepper (1987)
- Usability (section 6.2.3) – as suggested by Grice (2000)

These theories were not specifically developed to address the relationship between affect and design. However, researchers have identified attributes of these theories that may explain – at least in part – the relationship between affect and design. Each of these theories is unique in its approach and application to interface design. To enter into ‘flow’ certain conditions must exist in the environment and by entering this state a positive affective experience will be occur. The heuristics for internally motivating interfaces identifies specific system design characteristics that can be used to make people want to use a product, which subsequently will increase their affective experiences. Usability looks at the role of product behaviour upon the user, and if the
product behaves in a manner consistent with the user’s expectations or wishes, then a positive affective experience is more probable.

Despite these theories being unique in the way they attempt to understand the relationship between affect and design, there are some similarities. A comparison of the theories and areas of similarity are described in section 6.2.4. While comparing and contrasting the theories can provide insight into how they relate to each other, it does not begin to relate these theories to affect. It is tempting to suggest that the complementary elements of the theories are the ones that can be expected to create positive affect; however, this suggestion needs to be confirmed. For this reason, a study was designed to explore the relationships between the theories and affect, and is presented in section 6.3.

6.2.1. Flow

Csikszentmihaly describes flow as ‘the holistic sensation that people feel when they act with total involvement’ (Csikszentmihaly, 1975). In the state of flow, actions flow without cognitive intervention by the person. The term “flow” was used because people in this state often said that they “were in the flow of [the activity]”. With respect to affect and design, Draper (1999) suggested that the characteristics of situations that create flow may explain the appeal of computer games and other products that are intended to create a positive user-experience.

The original studies by Csikszentmihaly surveyed modern dancers, chess players, rock-climbers, classical composers, and basketball players – each of whom described the state of flow similarly despite the activities being vastly different (Csikszentmihaly, 1975). These studies began to look at why people engaged in activities that offered no reward except for the activity itself. In the first study, he asked participants from these groups to rank reasons for enjoying these activities. The participants consistently answered that the experience and the use and development of skills and the activity itself (pattern, action, and the context in which the activity takes place) were the primary reasons for engaging in an activity. In short, Csikszentmihaly found that people engaged in these activities because they enjoyed the internal rewards that the activity offered, and did not see external rewards as a vital aspect to participate in an activity.
With further research into flow, Csikszentmihaly defined the characteristics of flow-inducing activities (Csikszentmihaly, 1990). These are

The user:

- must feel capable of completing the task;
- must have the ability to concentrate on task;
- clearly recognises the goals of the task;
- receives immediate feedback about task performance;
- is removed from the awareness of worries and frustrations of the external world;
- has a sense of control over their actions;
- loses the awareness of themselves, yet has a stronger self-image after the activity; and
- has the sense of time altered: hours can seem like minutes.

These elements, combined in a single activity, are said to cause a sense of deep enjoyment so rewarding that people are willing to expend a great deal of energy simply to feel it. According to Csikszentmihaly (1975), flow can manifest itself in any situation or context that contains the above characteristics. It has subsequently been suggested that human-machine systems can create flow in the user and may explain why certain products will elicit fun and other positive affective responses in a user (Draper, 1999).

### 6.2.2. Internally motivating interfaces

Malone (1983), in agreement with Csikszentmihaly, believes that fun and enjoyment only arise from activities that people internally want to do. While Malone does accept that external rewards may enhance any positive affective experience, he believes that if the activity is to create a positive experience then the overriding reason for participating must be internally motivated.

Expanding upon his earlier work, Malone with the support of Lepper (1987) developed seven heuristics for the design of intrinsically motivating interfaces for instructional environments. They believed that if these characteristics are included in the design of
an educational system that the user will want to learn and actually enjoy learning. One identified caveat is that the motivational elements should work in conjunction with and not in opposition to the system’s instructional capabilities. This brings up the concept of trade-offs between requirements (see section 2.4.2).

Of the seven heuristics, four were considered ‘major’ factors related to intrinsic motivation and three were considered ‘minor’ factors related to external motivators.

The four major heuristics are:

1. **Challenge** – The interface should have multi-layers of challenge so that the user will feel initial success, and continue to see improvements (sense of accomplishment)

2. **Curiosity** – The interface should lead users to believe that their knowledge structures (or skills) are incomplete or inconsistent, and make the user want to strive to solidify these

3. **Control** – The interface should make the user feel that the outcomes are determined by the users own actions

4. **Fantasy** – The interface should evoke mental images of physical or social situations not actually present, and should be designed to the emotional needs of the user and so that the user can identify with the characters within the interface

The three minor factors include:

1. **Competition** – The interface should provide the user with some method of comparing their skills with those of other users or to benchmarks set within the system

2. **Cooperation** – The interface should allow the user the opportunity to work with others (system or other users) to promote interactive learning in a social environment

3. **Recognition** – The interface should allow the user to recognize the purpose of the interface elements presented to them

The minor factors were included because Malone and Lepper (1987) recognised that they may enhance the positive affective experience, however they are unlikely to create a positive affective experience in the absence of the major factors. This is congruent with motivation research where it has been said that internally motivated activities will
provide a more positive and meaningful experience than activities based on external motivators (Brown, 1988).

6.2.3. Usability

In ISO 9241-11 (1998), usability is characterised by three elements: effectiveness, efficiency, and satisfaction. These three elements and their impacts upon system quality were discussed in more detail in section 2.6.3. It should be reminded that usability as used within this thesis focuses on the underlying concepts, and not the processes commonly linked with usability. These processes (e.g. user-based testing, interviews, participative design) are considered part of human-centred design processes (ISO, 1999).

Grice (2000) assessed the affective experiences of users of computer games with respect to these three elements of usability. His hypothesis was that enjoyable computer games will have a high level of usability. Several minor experiments conducted by research students seemed to indicate that this hypothesis was true. However, the results of these studies seem to have been interpreted with bias towards accepting usability as a contributor of computer game enjoyment. There are many examples where the response of a participant was interpreted as relating to usability, whereas another observer would likely relate it to another quality. For example, in one study involving the use of an ice hockey game it was concluded that if non-hockey players were able to get more background information about the players on their team then they would find the game easier to learn, and therefore more enjoyable. However, a critique of this conclusion is that the context of the game (i.e. ice hockey) may be the reason the game is not being enjoyed regardless of the learnability. If a user has no interest in the context portrayed in the system, or no interest in the system itself, it is unlikely that they will find it enjoyable regardless of the usability.

Counterpoint to Grice (2000), Carroll and Thomas (1988) see fun and usability as incompatible. Their argument is that usability implies making things easy to use, whereas fun implies making things interesting and engaging – and this often requires an increase in the duration of product use and the integration of product design components that have an adverse impact upon usability. An example of this is provided by Nielsen
(1996) where he describes the ability of users to customise their desktops with various pictures and images. This ability decreases efficiency as it increases the underlying processing requirements of the computer in addition to the time people take to customise their desktops, and the sole benefit is the enjoyment of the user. However, Nielsen still states that usability and ease of use must exist before enjoyment of a system can occur.

In terms of usability and games, Federov (2002) states that satisfaction of the user is the prime usability consideration, and that effectiveness and efficiency are normally related to productivity of a system, which is normally not a concern to a computer game player. In a similar vein, Bickford (1997) says that the aim of a computer game is to keep the user engaged as long as possible – which is in direct contrast to the usability aim of efficiency. On the other hand, it is necessary that the player continues to see progress, and if the amount of effort they put forth does not equate to reasonable progress then they may become overly frustrated and no longer enjoy the game (Laird, 2005). This latter comment is heavily related to Malone and Lepper’s heuristic of ‘challenge’ (see 6.2.2).

Although there is some disagreement regarding the relationship between usability and affect, there are assertions that it can contribute to the affective experiences of the user. As such, it was still considered a viable theory relating design to affect.

### 6.2.4. Relationship between the Theories

Each theory addresses the relationship between design and affect from a different perspective. Flow tends to focus on the context of the situation, the heuristics tend to concentrate on elements of design in the product, and usability tends to focus on the behaviour of the product during use.

However, some similarities can be noted (see Figure 6.1). Flow and internally motivating heuristics have several common elements. This suggests that a design that focuses on internally motivating factors is also likely to increase user flow. Usability and flow also show some overlap, but there are also some elements that are incompatible. For example, the usability element of ‘efficiency’ is at direct odds with the flow element of ‘capable but not easy’ and the heuristic of ‘challenging’ (shown on
Figure 6.1 by use of a cross on the joining line). This suggests that a system designed with usability in mind may limit the degree of flow that a user can experience (and possibly limit the degree of affect — contrary to research by Grice, 2000).

6.3. **STUDY: Testing the Theories Relating Affect and Design**

This study was published in Bentley, Johnston, and von Baggo (2002c)

The theories of flow, heuristics for internally motivating interfaces, and usability have each have been alleged to address or provide insight into the relationship between affect and design (recall section 6.2). These theories offer three perspectives describing this relationship:

- Flow looks at the overall context of use of the system;
- The heuristics for internally motivating interfaces look at the design elements of the system and the user’s motivations; and
- Usability looks at the system's behaviour.

A comparison of these theories revealed that there was overlap between them, as well as some ‘incompatible’ elements (see 6.2.4).
A web-based study was developed to determine the extent that the theories describe the relationship between design and positive affective experience. A web-based study was chosen to facilitate the recruitment of a large heterogeneous sample with a minimal resource cost. Further, the web-based nature allowed data from participants to be directly entered into a database, effectively eliminating the resources required to transcribe and encode data.

Participants were asked to identify a computer game that they enjoy, and rate the degree to which the core elements of each theory contributed to their enjoyment. Participants were able to provide open feedback regarding the game and add further interface elements that they find contribute to their enjoyment. They were able to identify and rate up to three computer games that they enjoyed. Computer games were chosen for reasons discussed in section 6.1.

6.3.1. Purpose of the Study

This study aimed to address the question of which, if any, of the base elements (see Figure 6.1) of flow, heuristics for internally motivating interfaces, or usability contribute to user enjoyment? Knowing which elements contribute to enjoyment will provide designers with a greater insight into how they can design a system to elicit a positive affective experience.

It is unlikely that any single theory tested will adequately explain the relationship between affect and design. This is assumed because each theory views the relationship from a unique perspective (i.e. flow – context; heuristics – design; and usability – system behaviour) and it is anticipated that aspects of each theory / perspective will contribute to positive user affective experiences. Further, as each theory was not specifically developed to assist in understanding the relationship between affect and design, it is possible that some elements do not contribute to positive affective experiences, and may in fact deter from them.
6.3.2. Method

a. Procedure

Participants were recruited from online list-serves related to HCI or computer gaming. List-serve moderators were contacted requesting permission to distribute the study advertisement – no moderators refused. Following approval, an e-mail post containing a brief description and a link to the study website was circulated.

Potential participants arriving at the study website were provided with details of the activities to be performed should they participate, and were assured that the only data entered by themselves would be collected. This was done to alleviate fears of providing information over the internet and to re-assure anonymity during participation.

Upon accepting these arrangements, participants completed the survey (see c). They first answered generic demographic and background questions. Following this, participants elected to rate up to three computer games that they enjoy on the elements of the three theories. This rating was done on a five-point scale ranging from ‘Definitely Disagree’ to ‘Definitely Agree’. Participants were also provided with the opportunity to provide open feedback regarding the game.

b. Participants

As mentioned, participants were recruited from several e-mail list-serves relating to HCI or computer gaming. According to the subscriber information provided by each list-serve, approximately 5000 potential participants received the posting. It was impossible to discern if all e-mail addresses were valid. According to one of the moderators of a list-serve, less than ten percent of the subscribers actively participate in the list-serve. It can be assumed that some of the remaining ninety percent of subscriber’s e-mail addresses are no longer valid. It was impossible to discern how widely the posting was distributed beyond these initial postings to list-serves. It is known that over 60% of participants received the survey advertisement from friends or ‘other’ (Figure 6.4), suggesting that the call for participants was widely distributed beyond the original postings.

The survey was also administered to a control group consisting of 25 post-graduate students at an Australian university. This was to ensure that the internet population was
representative of a known population. Comparison of the internet and control populations showed no significant differences with respect to their reasons for enjoying computer games (see section 6.3.4.a for more detail on the comparison). As the two populations did not reveal any significant differences, they were combined for the remainder of the study.

There were 18 data sets that were identified as false. False data sets were identified by:

- Nonsense type entered into text fields;
- Feedback from participants stating that they were just testing the survey, and were not filling it out for real; and
- The selection of a single answer for all questions posed to the participant (i.e. the participant selected ‘definitely agree’ for all questions).

After these data sets were removed, there were a 303 participants recruited providing 409 unique data sets for analysis.

Participant demographics of age group, gender, and source of recruitment are discussed below. Comparisons are made with a comprehensive survey conducted in 2001 by the Interactive Digital Software Association (IDSA) providing a good representation of demographics of computer-game players. This provides a benchmark to which this study’s demographics can be compared.

**Participants by age group**

![Figure 6.2: Age group of participants](image-url)
The age breakdown of respondents does not conform to an extensive game playing habit survey conducted by the Interactive Digital Software Association (IDSA) where they found that 42% of people who play computer games are above the age of 35 (Interactive Digital Software Association, 2001). This present study indicated approximately 37% of participants were above the age of 30, and only 18% of participants above the age of 35. It is expected that the subscribers to the computer game list-serves targeted were primarily younger, and thus age was positively skewed in this study.

**Participants by Gender**

![Bar chart showing gender distribution of participants](image)

**Figure 6.3: Gender of Participants**

The IDSA survey found that 43% of computer game players are female (Interactive Digital Software Association, 2001), so the 18% female participation in this study was lower than expected. This disparity can be attributed to using computer game list-serves to recruit participants as females are less likely to join an online forum discussing games than are males.
Figure 6.4 represents how the participants were recruited. The purpose of this demographic was to indicate the representativeness of the sample population. If the participation was found to come from a single source (e.g. gaming list-serves) population then results may be biased towards the opinions of a select group. It was found that most participants heard about this survey through friends or ‘other’. The high selection of the ‘other’ category is probably due to the fact that a widely read website posted a link to the survey, unbeknownst to the researchers. This link was discovered when reading participant comments. However, the participant sourcing was not overly biased towards a single identifiable group.

**c. Materials Used**

A web-based survey was developed to assess people’s reasons to play computer games, based upon the three theories (see section 6.3.3 for survey pilot testing and development of the survey). The results of the survey were automatically submitted to a MySQL database. Security precautions against MySQL hacking were programmed into the survey.

The survey contained three main sections: Demographics; Reasons for playing computer games; and a Comments section. These sections are described below, and the complete survey can be viewed in Appendix C: .

**Section A: Demographics.** The aim of this section was to obtain some basic demographic information from the participants. Participants were asked to select their
age group, gender, occupation, computer game playing habits (number of games played regularly, and genre preference), and how they heard about the survey.

Section B: Reasons for playing computer games. This section asks the participant to identify a computer game that they enjoy, subjectively rate the enjoyment level, approximate the time that they spend playing this computer game, and then rate the theorised elements in regards to their contribution towards the user’s enjoyment, which was used to represent the user’s overall positive affect. Participants responded on a five point scale from highly agree to highly disagree, with the additional option of selecting ‘Not Applicable’ if the participant felt that the reason posed did not apply to the game they were rating.

Section C: Comments. An open-ended question was provided at the end of the survey. This allowed participants to provide comment on any aspect of the computer game that contributed to their enjoyment, but was not addressed elsewhere in the survey.

6.3.3. Survey Pilot Testing

Prior to administering the survey to the main group of participants, the survey was assessed by an acknowledged survey expert to ensure the questions were clear. Following this, the survey was tested with pilot participants. This involved the completion of a paper-based version of the survey followed by a focus group session to discuss the survey. The focus-group participants involved twelve individuals attending a weekly HCI research discussion group at an Australian University. The participants included academics, industry representatives, students, and general staff.

The original survey contained 15 elements corresponding to:

<table>
<thead>
<tr>
<th>Element</th>
<th>Theoretical Basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>Usability</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>Usability</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Usability</td>
</tr>
<tr>
<td>Concentration</td>
<td>Flow</td>
</tr>
<tr>
<td>Goals</td>
<td>Flow</td>
</tr>
<tr>
<td>Feedback</td>
<td>Flow</td>
</tr>
</tbody>
</table>
During the focus group session, several pilot participants indicated that learnability was an additional factor they considered to alter their enjoyment of a game, and was not included in any of the theoretical categories. It was therefore added to the final survey administered.

It was also suggested that competition with yourself (i.e. trying to beat your own high score) and competition with another person or computer have very different motivating effects on an individual. This was also true of recognition as it is possible to receive recognition from the computer or peers and again these are likely to have different effects on an individual. Therefore these two categories were split into ‘Competition – other’ and ‘Competition – self’, and ‘Recognition – computer’ and ‘Recognition – other’ respectively.

One focus group participant commented that he plays computer games because he is able associate with the game and the experiences it provides (e.g., a football player can associate with a football game). The focus group agreed that this can increase experience of positive affect when using a computer game, and was included in the survey under the title of attribution.

“Satisfaction” was also widely discussed during the focus group because several of the focus group participants interpreted “satisfaction” differently (corresponding to discussion in 4.3.6). From this discussion, it was foreseen that “satisfaction” could be interpreted by participants of the web-based study differently. As such, the decision was made to remove “satisfaction” from the web-based version of the survey because it

<table>
<thead>
<tr>
<th>Element</th>
<th>Theoretical Basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness of world / distraction</td>
<td>Flow</td>
</tr>
<tr>
<td>Perception of Self</td>
<td>Flow</td>
</tr>
<tr>
<td>Altered time</td>
<td>Flow</td>
</tr>
<tr>
<td>Challenge</td>
<td>Flow, Heuristics (Internal)</td>
</tr>
<tr>
<td>Control</td>
<td>Flow, Heuristics (Internal)</td>
</tr>
<tr>
<td>Curiosity</td>
<td>Heuristics (Internal)</td>
</tr>
<tr>
<td>Fantasy</td>
<td>Heuristics (Internal)</td>
</tr>
<tr>
<td>Competition</td>
<td>Heuristics (External)</td>
</tr>
<tr>
<td>Cooperation</td>
<td>Heuristics (External)</td>
</tr>
<tr>
<td>Recognition</td>
<td>Heuristics (External)</td>
</tr>
</tbody>
</table>
was not considered reliable. As such, it would be impossible to draw any valid conclusions based on results of the satisfaction question because it would not be clear as to which interpretation of satisfaction the participant was using.

The resulting survey had nineteen distinct elements that were randomly arranged on the survey. These are presented in Table 6.2, which also notes the theoretical basis of the element and the comment that participants rated on the survey.

<p>| Table 6.2: Survey Elements, their Theoretical Bases, and Related Survey Questions |
|-----------------|-----------------|-----------------|
| <strong>Element</strong>     | <strong>Theoretical Basis</strong> | <strong>Related Question</strong> |
| Efficiency      | Usability        | I enjoy this computer game because I can quickly figure out how to achieve the objectives of the game or level. |
| Effectiveness   | Usability        | I enjoy this computer game because I can consistently complete the game or levels with few errors. |
| Concentration   | Flow             | I enjoy this game because the information presented is not confusing. |
| Goals           | Flow             | I enjoy this computer game because it is clear what must be achieved to complete a level or win the game. |
| Feedback        | Flow             | I enjoy this computer game because the computer responds appropriately to actions taken by others, the game, or myself. |
| Distraction     | Flow             | I enjoy this computer game because it makes me forget about my worries and frustrations while I am playing. |
| Perception of Self | Flow           | I enjoy this computer game because after I stop playing it I feel better about myself. |
| Altered time    | Flow             | I enjoy this computer game because it makes time seem to go faster - hours can pass by in minutes. |
| Challenge       | Flow, Heuristics (Internal) | I enjoy this computer game because it is challenging, but within my ability to complete. |</p>
<table>
<thead>
<tr>
<th>Element</th>
<th>Theoretical Basis</th>
<th>Related Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Flow, Heuristics (Internal)</td>
<td>I enjoy this computer game because it lets me do what I want when I want.</td>
</tr>
<tr>
<td>Curiosity</td>
<td>Heuristics (Internal)</td>
<td>I enjoy this computer game because I am always wondering what will happen next, or what will occur if I take a specific action.</td>
</tr>
<tr>
<td>Fantasy</td>
<td>Heuristics (Internal)</td>
<td>I enjoy being able to do things in this computer game that I cannot do in everyday life.</td>
</tr>
<tr>
<td>Competition - Self</td>
<td>Heuristics (External)</td>
<td>I enjoy this computer game because I can try to beat my personal bests.</td>
</tr>
<tr>
<td>Competition – Other</td>
<td>Heuristics (External)</td>
<td>I enjoy this computer game because I can compete with others who have played it by trying to beat them or their scores.</td>
</tr>
<tr>
<td>Cooperation</td>
<td>Heuristics (External)</td>
<td>I enjoy playing this game because I can play it with others.</td>
</tr>
<tr>
<td>Recognition - Other</td>
<td>Heuristics (External)</td>
<td>I enjoy getting recognition from other people when I play the computer game well.</td>
</tr>
<tr>
<td>Recognition - Computer</td>
<td>Heuristics (External)</td>
<td>I enjoy the recognition I get from the computer when I play the computer game well.</td>
</tr>
<tr>
<td>Attribution</td>
<td>(Focus Group)</td>
<td>I enjoy this computer game because I can identify with the situations and/or people presented in this computer game.</td>
</tr>
<tr>
<td>Learnability</td>
<td>(Focus Group)</td>
<td>I enjoy this computer game because I do not need to devote much time and energy to learn how to play.</td>
</tr>
</tbody>
</table>

6.3.4. Data Reliability Checks

Statistical analysis was performed using SPSS v. 11.5. To ensure data reliability several checks were performed to test for the impact of potential extraneous variables on the
results. These included inter-participant reliability and demographic checks based on age, gender, and source of participant.

a. **Inter-Group Reliability**

There were two groups of participants:

- Control Group: post-graduate students at an Australian University
- Internet Group: participants recruited from e-mail list-serves

The use of the two independent groups was to ensure that each group provided similar results in terms of the reasons to play games. If results differed significantly the results of the survey may not be representative of the wider population.

A one-way between-groups multivariate analysis of variance (MANOVA) was performed to identify any difference between the control and internet populations and their reasons for playing games. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted.

There were no significant differences noted between the control and the internet sourced populations \( F(19, 406) = 0.564, p > 0.05 \), providing support that the internet sourced population was a representative sample.

b. **Demographics Reliability Testing**

As discussed in section 6.3.2.b (page 104), there seemed to be a disparity between groups participating in this survey and the survey conducted by the IDSA (2001) in regards to group demographics. The IDSA survey targeted thousands of households in the United States of America, and has been considered representative of computer game players worldwide, whereas this survey had a sample of 303 participants. To alleviate possible participant recruitment bias comparisons of age, gender, and source of participant were conducted. The demographic variables were used as independent variables to categorise participants to see if a specific demographic grouping influenced their reasons to play games. If there are no differences within these demographics then the results can be generalised to a wider population.
One-way between groups MANOVAs were performed on each demographic variable to identify if it had any impact upon reasons for playing games. Assumption testing was conducted for analysis to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted.

**Age**

It was found that age also did not influence the reasons why people play specific genres of computer games \( F(95, 860)=0.315, p>0.05 \). There was a trend for older participants to play games that were more individual and mentally active than younger players (i.e. simulation games or puzzle style games). Younger players tended towards the more active games (i.e. action/combat). However, this trend did not influence the actual reasons for enjoying the game. Interpreted, this means that an older game player playing a specific genre of computer game chooses to play it for the same reasons as a younger player playing the same genre. This suggests that as people age their actual reasons to play games may change towards the individual and mentally active style games. It also suggests that certain games will fulfil certain needs within a user, and users will select games that most appropriately meet their needs.

**Gender**

The MANOVA revealed no significant differences between gender and their reasons for enjoying a computer game \( F(19,188)=0.917, p>0.05 \). There were no observable trends in regards to the genres of computer games that males and females play.

**Source of Participant**

The source of the participant – how they were recruited – did not have any significant differences noted with respect to their reasons for enjoying a game \( F(95, 855)=1.079, p>0.05 \). This means that the variety of sources from which the participants were recruited provided a relatively homogeneous population in terms of their reasons for playing computer games. There were a large number of participants who stated that they were referred to the survey by friends or ‘other’ (see Figure 6.4) and this may have
a moderating effect as it is not possible to make assumptions about the population they came from.

6.3.5. Results and Interpretation

All data analysis was performed using SPSS v. 11.5

a. Theory Reconstruction

Principal components analysis (PCA) was used to reconstruct the theories used in the development of this survey. As there were three unique theories that were used to develop the survey it was expected that there would be three resulting components – one representing each theory.

Prior to performing the PCA, the suitability of data for factor analysis was assessed. Inspection of the correlation matrix revealed the presence of many coefficients of 0.3 and above. The Kaiser-Meyer-Oklin value was 0.804, exceeding the recommended value of 0.6 (Kaiser, 1970) and the Bartlett’s Test of Sphericity (Bartlett, 1954) reached statistical significance, supporting the factorability of the correlation matrix. Combined, these suggest that the use of PCA is acceptable for this data.

The nineteen reasons for enjoying a computer game were then subjected to PCA. The PCA was first performed without restrictions on the number of components, allowing the results to guide further processing of the data. Initial inspection of the scree plot generated by SPSS (see Figure 6.5) indicated that there were four components. The PCA revealed the presence of four components with eigenvalues exceeding 1, explaining 23.7%, 13.6%, 10.9%, and 6.2% of the variance respectively.
Based on this result, PCA was performed again restricting the number of components to four. Varimax rotation was performed to aid in the interpretation of results. The rotated solution is presented in Table 6.3 and reveals the presence of simple structure, with the four components showing strong loadings, and most elements loading substantially on only one component. The final four-factor solution explained a total of 54.54% of the variance, with components contributing 15.7%, 14.7%, 12.1%, and 12.1% respectively.
Table 6.3: Final Rotated Component Matrix

<table>
<thead>
<tr>
<th>Reason Element</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
<th>Component 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competition - Self</td>
<td>.883</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperation</td>
<td></td>
<td>.826</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition - Computer</td>
<td></td>
<td>.814</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer recognition</td>
<td></td>
<td>.538</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competition - Other</td>
<td></td>
<td></td>
<td>.542</td>
<td></td>
</tr>
<tr>
<td>Goals</td>
<td></td>
<td></td>
<td>.771</td>
<td></td>
</tr>
<tr>
<td>Concentrate</td>
<td></td>
<td>.729</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>.734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learnability</td>
<td></td>
<td>.729</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness</td>
<td></td>
<td>.492</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td></td>
<td></td>
<td>.674</td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td></td>
<td>.595</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curiosity</td>
<td></td>
<td>.606</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
<td>.638</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribution</td>
<td></td>
<td>.534</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>.424</td>
<td></td>
<td>.539</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td>.684</td>
<td></td>
</tr>
<tr>
<td>Self Awareness</td>
<td></td>
<td></td>
<td>.693</td>
<td></td>
</tr>
<tr>
<td>Distraction</td>
<td></td>
<td></td>
<td></td>
<td>.598</td>
</tr>
</tbody>
</table>

The results of this analysis identified four components that did not align with the three theories initially used to develop the survey. However, upon inspection of the rotated matrix it revealed that the fourth and ‘unexpected’ component was caused by the division of internal and external motivating heuristics that were originally part of a single theory. The four categories identified and their respective components are shown in Table 6.4.
**Table 6.4: Resulting Factors from the Principal Components Analysis**

<table>
<thead>
<tr>
<th>External Motivators</th>
<th>Usability</th>
<th>Internal Motivators</th>
<th>Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competition – Self</td>
<td>Goals</td>
<td>Challenge</td>
<td>Control</td>
</tr>
<tr>
<td>Cooperation</td>
<td>Concentration</td>
<td>Fantasy</td>
<td>Time</td>
</tr>
<tr>
<td>Recognition - Computer</td>
<td>Efficiency</td>
<td>Curiosity</td>
<td>Self Awareness</td>
</tr>
<tr>
<td>Peer recognition</td>
<td>Learnability</td>
<td>Feedback</td>
<td>Distraction</td>
</tr>
<tr>
<td>Competition – Other</td>
<td>Effectiveness</td>
<td>Attribution</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td></td>
</tr>
</tbody>
</table>

The only element that was found to exist within more than one component was control, and this was found to exist within both the internal motivators and within flow. Control was previously recognised as being a component in both these theories (see Table 6.2) and was not unexpected. It was also found that the goals and concentration elements were contained in the component of usability rather than flow. While not anticipated, it is accepted that clear goals are commonly associated with high usability. For example, having clear goals is the first step when performing a cognitive walkthrough of a design (Polson et al., 1992). It was, however, unexpected to find concentration to be a part of usability rather than flow.

The net result from the PCA is that the components tended to align well with the theories that were used initially. The only notable exception is that some elements of flow were better described as elements of usability within the context of interactive systems. This is acceptable because it was already known that flow and usability did contain some overlap (see Figure 6.1).

**b. Predictors for Affect**

Standard multiple linear regression was used to determine which of the 19 reasons rated within the survey most greatly predicted the enjoyment of computer games. Assumption testing for the use of multiple linear regression was conducted which included testing for multicollinearity, outliers, and homoscedascedity. All assumptions were met.

The independent variable used for the multiple regression was enjoyment rating, and the dependant variables were the 19 individual factors. Results of the regression are shown in Table 6.5. Shaded variables are significant at the p<0.05 level.
Table 6.5: Results of standard linear regression

<table>
<thead>
<tr>
<th></th>
<th>Unstandardised Beta</th>
<th>Standardised Beta</th>
<th>Sig.</th>
<th>Summary Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>.452</td>
<td></td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Ability to Concentrate</td>
<td>.005</td>
<td>.013</td>
<td>.808</td>
<td></td>
</tr>
<tr>
<td>Computer Competition</td>
<td>-.010</td>
<td>-.031</td>
<td>.542</td>
<td></td>
</tr>
<tr>
<td>Clear Goals</td>
<td>-.023</td>
<td>-.072</td>
<td>.203</td>
<td></td>
</tr>
<tr>
<td><strong>Learnability</strong></td>
<td><strong>.047</strong></td>
<td><strong>.166</strong></td>
<td><strong>.002</strong></td>
<td>Less learnable predicts increased enjoyment</td>
</tr>
<tr>
<td>Feedback</td>
<td>-.071</td>
<td>-.171</td>
<td>.001</td>
<td>More feedback predicts increased enjoyment</td>
</tr>
<tr>
<td>Distraction Element</td>
<td>-.021</td>
<td>-.059</td>
<td>.246</td>
<td></td>
</tr>
<tr>
<td>Control over actions</td>
<td>-.029</td>
<td>-.085</td>
<td>.113</td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>-.009</td>
<td>-.029</td>
<td>.614</td>
<td></td>
</tr>
<tr>
<td>Computer Recognition</td>
<td>.030</td>
<td>.104</td>
<td>.108</td>
<td></td>
</tr>
<tr>
<td><strong>Self Image</strong></td>
<td><strong>-.070</strong></td>
<td><strong>-.217</strong></td>
<td><strong>.000</strong></td>
<td>Increased self image after use predicts increased enjoyment</td>
</tr>
<tr>
<td>Self Competition</td>
<td>.014</td>
<td>.049</td>
<td>.461</td>
<td></td>
</tr>
<tr>
<td><strong>Loss of Time</strong></td>
<td><strong>.029</strong></td>
<td><strong>.101</strong></td>
<td><strong>.043</strong></td>
<td>Less loss of time predicts increased enjoyment</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>-.014</td>
<td>-.042</td>
<td>.396</td>
<td></td>
</tr>
<tr>
<td><strong>Curiosity</strong></td>
<td><strong>-.048</strong></td>
<td><strong>-.146</strong></td>
<td><strong>.004</strong></td>
<td>More curiosity predicts increased enjoyment</td>
</tr>
<tr>
<td>Attribution</td>
<td>.003</td>
<td>.010</td>
<td>.850</td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>-.019</td>
<td>-.049</td>
<td>.342</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>-.031</td>
<td>-.059</td>
<td>.228</td>
<td></td>
</tr>
<tr>
<td><strong>Cooperation</strong></td>
<td><strong>-.034</strong></td>
<td><strong>-.117</strong></td>
<td><strong>.047</strong></td>
<td>More cooperation predicts increased enjoyment</td>
</tr>
<tr>
<td>Peer Recognition</td>
<td>.000</td>
<td>.000</td>
<td>.999</td>
<td></td>
</tr>
</tbody>
</table>
The R-square value for this regression was .219. The variables that most greatly predicted computer game enjoyment were

- a reduced learnability,
- high feedback,
- increased perception of self,
- not losing time,
- increased amount of curiosity, and
- increased cooperation.

The results related to ‘reduced learnability’ and ‘not losing time’ were unexpected as the theories predicted an increase in any of the elements should lead to an increase in user enjoyment. Further analyses are required to explain these unexpected results, and are provided at the end of this section (e).

c. Genre Influences on Affective Predictors

Different genres of computer games may emphasize different theoretical elements contributing to the positive affective experiences enjoyed by the user. The genre classification scheme used was provided by the online game review and research website International Hobo (2001), and is shown in Figure 6.6. To supplement the genre classification the relative orientation towards motor or cognitive skills has been added.

![Computer Game Genres and Relative Skill Orientation](image)

To test for differences between genres and the reasons for playing them a one-way between-groups MANOVA was performed. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers,
homogeneity of variance-covariance matrices, and multicollinearity. It was noted that equality of variance-covariance matrices was violated [Box's $F(380)=0.001$], however it has been said that this test can be too strict when using large sample sizes (Tabachnick and Fidell, 1996, p. 81). It was realised that several of the genres violated the assumption of equality of variances. These violations do not prohibit the use of a MANOVA, however it does mean that increased discretion is required to interpret the results. Therefore, the alpha level was set to a more conservative 0.01 rather than 0.05.

It was found that there was a statistically significant difference between genres on the reasons people enjoyed the game [$F(152, 3076)=2.152$, $p<0.01$, $\eta^2=0.089$]. Post-hoc analysis using a Bonferroni adjusted alpha level of 0.002 revealed significant differences for genre within learnability, control, curiosity and fantasy [statistical ranges from $F(8, 431)=4.494$ to $7.187$, $p<0.002$]. Significant results are summarised in Table 6.6.

<table>
<thead>
<tr>
<th>Table 6.6: Post-Hoc Results for Genre Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td><strong>Strategy</strong></td>
</tr>
<tr>
<td><strong>Individual (Puzzle)</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

It can be noted that individual style games differed from strategic, role playing, action / combat, and racing / flying games for several different reasons. Individual style games tend to be based on existing real-world puzzles and games and therefore do not promote curiosity or fantasy, as people already know what to expect.

Control also did not contribute to the enjoyment of individual games as much as it did with others. Presumably this is because control in an individual game tends to be a given – that is the game rules prescribe what can and cannot be done at a given time, and therefore control is a non-issue for these games. Other game genres give the user greater freedom to choose various actions at any given time.
In contrast to other genres, learnability in individual style games was considered very important (see Figure 6.7). Personal experience suggests that individual and puzzle style games tend to be played as a distracter or during bouts of procrastination. At these times, users are unlikely to want to devote too many resources to learning the game, instead wanting to just play. Learnability is further discussed in section e.

![Learnability by Genre](image)

Figure 6.7: Learnability by Genre

These results suggest that certain system types will draw upon different ‘inspirations’ for user enjoyment. Therefore, the context of use and nature of the system must be considered with regards to the choice of design solutions if a positive affective experience is desired. Simply including the elements described within this study in a design will not guarantee positive affect or user enjoyment. Knowledge of what the user wants is important in the appropriate selection of design elements.

d. Survey Comments

The survey comments section revealed that the theories do not address all reasons for the enjoyment of modern computer games. Groupings of comments were found to relate to aesthetics, online competition, and ‘distracter’ games. All comments cited were spell-checked, and no other alterations were made – emphasis and punctuation marks are those of the participant.
**Aesthetics**

Many participants shared a reference to the aesthetic appeal of computer games. Some sample comments are as follows:

- It is graphically very appealing.
- I enjoy the game because of the graphics, physics, level and map design.
- Looks great!
- Love the cheesy voice-overs and acting.
- I enjoy the game because of its highly cinematic effects, its cleverly designed space and the innovative take on horror film conventions.

It should be noted that flow and the heuristics for designing intrinsically motivating interfaces were each developed before quality graphics became available in interactive systems. In addition, the concept of usability was developed primarily for production contexts, in which aesthetics is considered a desirable attribute but not a need (especially if it detracts from usability). As such, the three theories did not include aesthetics.

Crawford (2004) has labelled aesthetics as a marginal quality in relation to computer games, meaning that it is not a prime quality that will ‘make or break’ a system. According to Crawford, marginal qualities, such as aesthetics, influence the systems quality when all prime qualities are similar between systems. Lindgaard (2001) agrees with this saying that most mobile phones are equal in terms of functionality and the services offered (prime qualities), and often it is the aesthetic appeal of the phone that will influence a person’s decision to purchase it.

While it is understood why the three theories neglected aesthetics the comments made by participants indicate that quality aesthetics and graphics can influence positive affective experiences. However, aesthetics should be considered only after other elements are considered. Aesthetics may increase user enjoyment but it will not, by itself, create a positive affective experience. It must be supported by other ‘prime qualities’ that promote an enjoyable experience. This is captured in one user’s comment: ‘This is an OLD game. Made in 1987. Classic RPG. Looks bad, sounds bad, but IS damn good.’
Online Competition

Many participants commented on the ability to play and compete against other human players online. Malone and Lepper’s (1987) heuristics address competition, and the survey expanded this to include both competition against self (i.e. high scores) and competition against another player or agent. The survey did not distinguish between playing another human player versus playing against an artificial computer opponent. Based on participant comments, this is a limitation to the survey as they did see a distinction:

- I primarily play in multi-player mode, against real people. Live players make the game much more interesting than the current AI available.
- This game is really cool because each person has their personal way to play, making battles instable; you don’t know what will happen, when it will happen. You can make more defence or attack (it’s the better defence). However sometimes people disagree with this theory (which is why it is interesting).
- It is very cool try to win against other players counting on only your wit and intelligence.
- I enjoy it because I’m not playing against an AI – I am playing against another player from across the globe.

Therefore competition should be considered on three different aspects – competition against yourself, competition against the computer, and competition against another player (online or in person). The comments provided suggest enjoyment increases when playing against another human player, and the survey results did not reveal this because the question did not distinguish between playing a computer versus playing another user through the computer.

Distracter Games

For individual and puzzle games (e.g. Tetris, solitaire), participants often made the comment that they are used as a distracter. Comments to this effect included:

- It’s a good time waster
- It is pleasantly mind-numbing!
• I play it when waiting or wasting time when my thesis is getting me down. It's an avoidance thing. Sometimes I feel bad after I have played too long.

These comments relate to the desire of the user – they want to do something that will temporarily take their mind off their current tasks or to procrastinate. This again emphasises that the enjoyment of the game will vary based upon the context in which the game is played. Individual and puzzle games seem to be well suited for short bursts of playing, whereas some other genres require extended playing time to get the sense of enjoyment from them.

e. **Explanation of Unexpected Results**

As mentioned in section b unexpected results were found for the elements of learnability and loss of time. It was found that decreased learnability increased game enjoyment, and that the less loss of time increased enjoyment. According to the theories, increases in each of these should have increased the positive experience. Further interpretation of these two unexpected results was left until now because the additional analyses performed in c and d can aid in the explanation.

**Learnability**

It was found that learnability is a benefit to individual and puzzle games, but was a detriment to other game genres. Individual and puzzle games tend to be used as distracters, and users do not wish to devote resources to learning how to play – they simply want to play. Other genres see learnability as a detriment though. This is because learning how to play is part of the challenge of the game. Learning new strategies or learning how to better control the user’s character in the game is what gives the user enjoyment.

**Loss of Time**

For all computer game genres a decreased amount of ‘loss of time’ was strongly related to greater computer game enjoyment. This is converse to the flow theory, which suggests that ‘loss of time’ was indicative of flow and therefore positive affect. Explanation of this unexpected result can be made by analysing the statement that the participants responded to, which was *I enjoy this computer game because it makes...*
time seem to go faster - *hours can pass by in minutes*. This statement implies that the user enjoys losing time. However, the results show that losing time is not something that the game player enjoys; rather it is a consequence of playing a game that they do enjoy. Therefore losing time is an indicator of flow and positive affective experience – it does not create the experiences.

### 6.3.6. Conclusions

The aim of this study was to discern which elements of flow, heuristics for internally motivating characteristics, and/or usability can increase enjoyment. By understanding what elements increase enjoyment it is possible to target these elements during the design stages with the desired outcome of enhancing the user’s enjoyment. The study revealed that there were six elements of these theories that predicted a greater amount of enjoyment. These were:

- Learnability
- Feedback
- Curiosity
- Cooperation
- Self-image
- Loss of time

These elements of the computer games predicted the greatest impact upon enjoyable experiences. Participants were also provided the opportunity to comment on other aspects of the system that enhanced positive affective experiences. Many participants commented on two areas, which were:

- Aesthetic appeal
- Competition against other human players

Reviewing the eight elements listed some are designable and tangible aspects of system interaction (i.e. learnability, feedback, curiosity, cooperation, aesthetics, human competition), whereas others are internal to the user (i.e. self-image, loss of time) and are not tangible in design. These are addressed in 6.3.6.a and 6.3.6.b respectively.
a. Designable Elements Related to Affect

The results from the study make it possible to create a set of design guidelines that will increase the enjoyment experienced by a user of a system. Table 6.7 provides this set of guidelines.

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Ensure that the user has adequate amounts of feedback to know what the system is doing, and that it is responding appropriately to user input.</td>
<td>Statistical result: 6.3.5.b</td>
</tr>
<tr>
<td>• Keep the user curious as to what will happen next – do not make the system transparent in such a manner that the user always knows what will happen and what to do.</td>
<td>Statistical result: 6.3.5.b</td>
</tr>
<tr>
<td>• Increase the aesthetic appeal of the interface by including high quality graphics, sounds, and images appropriate to the design.</td>
<td>Comments: 6.3.5.d</td>
</tr>
</tbody>
</table>
| • Vary the learnability of the system with respect to the context of use of the system.  
  o Ensure a high level of learnability if the system is to be used for short periods of time or to fill in time (i.e. while waiting).  
  o A lower level of learnability can be advised when success for the user is based upon their learning of strategies or learning how to best control / manipulate the system. | Statistical result: 6.3.5.e |
| • Increase the amount of cooperation required between users, increasing social interaction where each person is striving to achieve a common goal. | Statistical result: 6.3.5.b |
| • Promote competition against other humans where feasible.                 | Comments: 6.3.5.d           |

The final two guidelines seem to be opposed, with one suggesting cooperation and the other suggesting competition. This is not always true – it is possible to use ‘team play’ in which one team competes against another team. Thus, the user is cooperating with his or her team, while competing with another. However, it is not always necessary or possible to implement each guideline for every system. It is the responsibility of the designer to select which guidelines will enhance the enjoyment of the user.

Each guideline must be considered within the context of use for the system and be implemented appropriately. For instance, increasing the amount of feedback beyond
what is required for the user to know what is happening or decreasing learnability to a degree where the system is unable to be used will result in negative experiences.

These guidelines are quite similar to other guidelines recently developed in relation to computer games and design. Adequate (and appropriate) feedback to the user has been suggested to be a critical element in game design by Laird (2005) and Bickford (1997), Laitinen (2005; 2006) in addition to Malone and Lepper (1987). ‘Communication of status’ has also been identified as a class of computer game design patterns (Folmer, 2006), and is strongly related to feedback. Learnability is another design pattern described by Folmer (2006), and was also identified by Grice (2000) as a key contributor to game success. Finally Aesthetic appeal is often linked with quality of games: Laird (2005), Bickford (1997), Grice (2000) – though a commonly cited caveat is that aesthetics is only to enhance overall experience and cannot create the experience (Norman, 1988). It is interesting to note that most of the guidelines provided in the literature are based experience or anecdotal evidence, and that this study provides a degree of validation of these.

Although these guidelines show significant similarity with other research, there were some guidelines identified that did not relate with other work. Significantly, neither cooperation nor human competition is identified in other game design guidance. However, this may simply be due to the nature of the research reviewed – most relates to a single player and single game. It is expected that as researchers begin to look at multiplayer online games that human cooperation and competition will become more prominent in design guidance. Indeed, Smith (2004, 2005a, 2005b) has recently described how other players in a multiplayer game may influence the overall enjoyment – both in terms of cooperation and competition – though refrained from giving design guidance, instead focussing on the social interactions. Smith is continuing research in this area, and the reader is encouraged to read the latest findings at Smith (2006) if interested.

As with most system requirements, the selection and prioritisation of the guidelines in Table 6.7 is subject to trade-offs with other requirements. If the system relies heavily on the affective experiences of users then it may be appropriate to address most of the guidelines. For performance-based technologies, these design elements may detract
from the overall productivity of the user, and must be chosen accordingly. Trade-offs and requirements was discussed in greater detail in section 2.4.2.

b. **Non-Designable Elements Related to Affect**

Non-designable elements that were revealed as predictors of affect can only be observed during or after a user has interacted with the system – it is not possible to include them in the design. However, they can provide an indicator whether the design can produce a positive affective experience and may be give rise to evaluation measures. The non-designable elements revealed were:

- Self-image: Does the user have an improved self-image?
- Loss of time: Has the user ‘lost time’ during their system interaction?

While these seem to be simple evaluation strategies for affect, difficulties arise when trying to conceive of objective measures. For example, asking the person if their self-image has improved or how long they had been using a system for requires a subjective response, and it would be unknown whether the system or an external variable caused the result – it is possible that the user simply answered the way they thought the researcher or experimenter wished them to. Further, asking the user a simple question does not identify what design element or combination of elements that created the affective experience. Related to this, the affective experience is transient and asking the user a question at the end of an interactive session is better representative of the overall emotion rather than specific affective events. Further, it is not possible to ask questions during system use for fear that it will alter behaviour and change the affective experience (Fishbein and Ajzen, 1975; Omodei et al, 1994) and therefore any answer will not truly represent the affect experienced in a non-interrupted session. It can be assumed that if the user had an overall positive emotional experience that there were positive affective experiences as well, but again it is not possible to discern which aspects of the system created the individual affective experiences.

The issues of developing suitable validating measures for affect are significant. Key to the development of affect validation is that it must be viable, must not interrupt or change the user-experience, and ideally should be objective.
6.4. Chapter Summary

Accepting that affect is an important element of QIU is insufficient – there must also be design guidance that links affect and product design. Without this knowledge, designers may neglect affect as an element of product quality because it is not known how to address it within the design of a product.

This chapter began initial exploration into the holistic relationship between affect and design, and it offered some guidance to designers wishing to design a system that will create positive affective experiences, in particular enjoyment. In detail, this chapter:

- Highlighted issues in relation to studying the holistic relationship between affect and design. The primary issue was the need to select and use a product that does create affective experiences in users (i.e. testing a performance-based technology is not relevant). To this end, computer games were recommended as one type of product that will allow for the study of affect and design.
- Reviewed and contrasted three theories that were independently suggested as relating affect and design. These were flow, heuristics for internally motivating interfaces, and usability.
- Reported on a web-based survey that examined these three theories and their individual elements to identify what, if any, elements of these theories do provide insight into the relationship between affect and design. Results revealed that the learnability, feedback, curiosity, cooperation (with others), user self-image, user ‘lost time’, aesthetic appeal, and competition (with other individuals) were each associated with positive affective experiences.
- Provided design guidance based on the results of the study that relates design to the creation of positive affective experiences in users.

Aside from the design guidance, the results of the study also led to the discussion of affect evaluation methods. The discussion was brief, as the issues involved in the evaluation of affect necessitates a more detailed explanation that is beyond the scope and purpose of this chapter. Chapter 7 addresses these issues, and appraises current QIU evaluation methods for their ability to evaluate affect. Expanding upon this, Chapter 8 describes a study of several seemingly viable affect evaluation methods.
Chapter 7 Validation of Affective Requirements

The concept of affect, its relationship to system design, and why it is important to include it as a component of system quality has been the focus of this thesis thus far. It has been argued that affect is a prime predictor of QIU for non-performance-based technologies, and it may influence the QIU for some performance-based technologies.

However, accepting the importance of affect and knowledge of design elements that can contribute to affect is only one step towards the inclusion of affect within the concept of system quality. Simply integrating design characteristics known to aid in the development of user affective experiences (see section 6.3.6) does not guarantee that the affective requirements will be met – there is a need to validate that these requirements have been met by the system.

It is possible to validate the external quality of a product in relation to affect. This would involve determining whether a product has integrated design characteristics intended to create or enhance a user’s affective experience. However, having high external quality will only influence QIU, and does not necessarily mean that a product will have high QIU. Thus, even if the design characteristics related to affect have been integrated into a product it does not guarantee that the user will experience the desired affect.

It not necessary to validate every possible element of QIU within a system (recall Figure 2.4, and affect), rather it is only necessary to validate those that have an impact upon the actual quality perceived by a user. In cases of pure performance-based technologies, affect may not be an important quality and validation unnecessary. Conversely, for non-performance-based technologies, the validation of affect may be of primary importance, and it may be possible to ignore the validation of many performance-based qualities. This relative weighting of affect-based versus performance-based factors contributing to QIU was shown in Figure 5.2 and described in 5.3.
Within this chapter, the capabilities of some current QIU validation methods are critically reviewed with respect to their ability to validate affect. Based upon the results of this review requirements for affect validation are established. The chapter continues by outlining two possible QIU validation methods that can target affect – physiological indicators and the cued-recall methodology. A study of these methods and their actual ability to validate affect is provided in Chapter 8.

7.1. Stimulus-Affect-Measurement

While many performance-based quality characteristics have clear metrics for measurement (e.g. efficiency and effectiveness), the issue of measuring affective requirements poses problems. The primary issue is that it is virtually impossible to have an objective measure of affect, and user affect is very difficult to subjectively assess. In the case of performance-oriented requirements there are tangible ‘real-world’ objective measures that are directly indicative of the requirement (e.g. efficiency=time). However, the same cannot be said of affect-based requirements.

A useful way to look at the current situation of affect validation is to use Quality Attribute Scenarios, as discussed by Bass et al. (2003). Bass et al. use these scenarios to encourage looking at the entire pathway that impacts upon a given requirement (see Figure 7.1).

![Figure 7.1: Quality Attribute Scenario](From Bass et al., 2003)

Each quality attribute scenario should focus on a specific element of quality, and describes how the quality component manifests itself within the system. There are six parts of a quality attribute scenario, as described in Table 7.1.
Table 7.1: Elements of a Generic Quality Attribute Scenario

<table>
<thead>
<tr>
<th>Scenario Part</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source of stimulus</td>
<td>This is some entity (a human, a computer system, or any other actuator) that generated the stimulus.</td>
</tr>
<tr>
<td>Stimulus</td>
<td>The stimulus is a condition that needs to be considered when it arrives at a system.</td>
</tr>
<tr>
<td>Environment</td>
<td>The stimulus occurs within certain conditions (i.e. contexts).</td>
</tr>
<tr>
<td>Artefact</td>
<td>Some artefact is stimulated. This may be the whole system, or some pieces of it.</td>
</tr>
<tr>
<td>Response and Response</td>
<td>The response is the activity undertaken after the arrival of the stimulus. This response should be measurable in some fashion so that the requirement can be tested.</td>
</tr>
<tr>
<td>Measure</td>
<td></td>
</tr>
</tbody>
</table>

Bass et al. (2003) provide example quality attribute scenarios describing the qualities of availability, modifiability, security, and testability. In their examples, the artefact relates to the performance of the system with respect to a specific quality.

These scenarios can be used for affect-based qualities, if it is accepted that the artefact can be within the human aspect of the overall system (rather than the product). A generic affect quality attribute scenario is shown in Figure 7.2, where the artefact would be the affective experience of a user. Each element of the scenario, as they apply to affect, is described in Table 7.2.

Figure 7.2: Affect Quality Attribute Scenario
Table 7.2: Parts of the Affect Quality Attribute Scenario

<table>
<thead>
<tr>
<th>Scenario Part</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source of stimulus</td>
<td>The stimulus is generated through user interaction (perhaps with the product or other users).</td>
</tr>
<tr>
<td>Stimulus</td>
<td>The stimulus can be related to any system property that can influence or create affective experiences. The web-based survey found that the following system properties can achieve this (from 6.3.6.a): Learnability Feedback Curiosity Aesthetic appeal Cooperation with other human users Competition against other human users Learnability, feedback, curiosity, and aesthetic appeal are examples where the source of the stimulus is derived through general system interaction. Cooperation and competition are examples where the source of the stimuli is from other system users.</td>
</tr>
<tr>
<td>Environment</td>
<td>The environment (or context of use) varies in which non-performance-based technologies will be used. It may be a personal environment (e.g. home), work settings (office), or it could be in a public environment (e.g. park). The environment may also be a special case, such as a training facility (e.g. haptic devices to train surgeons).</td>
</tr>
<tr>
<td>Artefact</td>
<td>The artefact is the user’s affective experience.</td>
</tr>
<tr>
<td>Response and</td>
<td>It is unclear how to ‘observe’ an affective experience in a user, and as such it is not clear how to measure this response. Some current quality in use evaluation methods, as they relate to affect, are discussed in section 7.2.</td>
</tr>
<tr>
<td>Response Measure</td>
<td></td>
</tr>
</tbody>
</table>

7.2. Quality in use Evaluation Methods and Affect

There are no existing evaluation methods described as being capable of validating user-affect (as it has been defined within this thesis) in the context of product use. This does not mean that current QIU evaluation methods cannot validate affect – it simply means that the current methods have not been applied to affect, but may be capable of validating affect.
This section describes and critically assesses several QIU evaluation methods as they may apply to affect validation. While the evaluation methods described are not exhaustive of the array of methods that exist for QIU evaluation, they provide a good cross-section of available methods and serve to highlight important considerations that must be addressed in the validation of affect. Dumas and Redish (1993) provide a more comprehensive review of evaluation methods.

The QIU evaluation methods selected for review have been grouped into two categories – during-use and post-use evaluation methods. As the category labels imply, during-use evaluation methods occur while the user is interacting with the system and post-use evaluation methods occur after the user is finished interacting with the system. During-use evaluation methods described include the think-aloud technique and direct observation (section 7.2.1). Post-use evaluation methods include unstructured interviews, structured interviews, and questionnaires (section 7.2.2). These methods were reviewed because they have been used to evaluate factors not directly performance-based – such as thoughts, decision processes, emotions, and user satisfaction. One specific questionnaire, The Software Usability Measurement Inventory (SUMI), does identify “affect” as a measured dimension and, as such, is discussed separately in section 7.2.3. However, as it will be shown, the concept of “affect” within the SUMI is better related to user emotion or overall satisfaction than affect (as the terms have been defined within this thesis).

Expert evaluation methods, such as heuristic evaluation (Nielsen, 2004) and cognitive walkthroughs (Polson, Lewis, Rieman, and Wharton, 1992), have not been reviewed because they have the ability to measure the stimulus, not the resulting artefact (i.e. user affect). While the stimulus does give rise to the artefact and subsequent response there are many other variables that could impact affect as well, and therefore the expert evaluation methods are of little predictive validity. Further, the individual nature of affect (as it depends on past experiences) limits the use of expert evaluation techniques that apply rules of thumb or general stereotyped response expectations based on product characteristics.

This review highlights the strengths and limitations of these methods with respect to affect validation. These formed the basis of the affect validation method requirements.
(section 7.3), which were then used to determine alternative evaluation methods that may be viable (section 7.4).

7.2.1. **During-Use Evaluations Methods**

Evaluation methods that occur while the user is interacting with the system provide the opportunity for the user to express their thoughts and/or feelings in situ. The two during-use evaluation methods that are considered are the think-aloud method and direct observation.

**a. Think-Aloud Evaluation Method**

The think-aloud technique requires the user to vocalise their thoughts and reasons for actions (Van Someren, Barnard, and Sandberg, 1994). In the context of affect, it would require the user to describe their affect throughout interaction with a product. These user descriptions can then be interpreted or quantified in some way – such as the categorisation of positive and negative affective statements. The quality attribute scenario shown in Figure 7.3 demonstrates the think-aloud technique as applied to affect.

![Figure 7.3: Affective Quality Attribute Scenario for the Think-Aloud Evaluation](image)

**Figure 7.3: Affective Quality Attribute Scenario for the Think-Aloud Evaluation**

While the speak-aloud method provides a response that is measurable and even quantifiable, there are some identified issues when applying this to affect. First, this method relies on participants to express their feelings with an evaluator whom they are likely to have only just met. Although there will be considerable individual variation, it is expected that the majority of users will be hesitant and not forthcoming in regards to the expression of their affective experiences. For this reason Van Someren et al. (1994) encourage the evaluators to be approachable and open to participants.
A second issue is that the think-aloud method requires the user to change their normal behaviour and perform an additional task – thinking aloud and expressing their affect. Fishbein and Ajzen (1975) have previously shown that changing behaviour can alter a person’s affective experiences. Therefore any affect experienced by the user during the think-aloud evaluation session may be influenced by the nature of the evaluation method, and cannot be considered a valid representation of the artefact as it would be experienced outside the evaluation. To put this into the context of the quality attribute scenario, this actually changes the environment of the system, and the stimulus may create a different artefact in this abnormal environment.

b. Direct observation

Direct observation of the user actually using the system is an often-neglected form of QIU evaluation, with evaluators normally choosing methods where they have an active role. While it does not allow for the insight into thoughts and reasoning of actions (as the think-aloud technique does), it does allow an evaluator to garner an unhindered view of a person using the system.

People are astute in noticing minor changes in body language and facial expressions in other individuals, and it forms part of our everyday communication techniques (Crystal, 2003). If a person is smiling, one can usually assume they are happy, or if a person is tense then they may be frustrated or angry. In relation to affect, it is possible for the evaluator to observe and record changes in the user’s body language, behaviour, or facial expressions as an indicator of an affective experience. The affective quality attribute scenario for this evaluation method is shown in Figure 7.4.

![Figure 7.4: Affective Quality Attribute Scenario for Direct Observation](image-url)
As opposed to the think-aloud method, direct observation does not require any additional tasks to be performed by the user. The Hawthorne effect, as interpreted by Mayo (1933), suggests that simply watching an individual will affect their behaviour and performance. However, Csikszentmihaly (1990) has shown that when an individual is deeply engaged that they will become oblivious to the activities around them as they are so immersed within their own activity. The Hawthorne studies were conducted in a factory where the workers are unlikely to be deeply highly engaged or immersed in their activities, and thus the act of observation may have influenced the behaviour of the workers. For affect research purposes, users will normally become deeply engaged because that will promote user affect, and therefore direct observation can be considered valid. It should also be noted that the true nature of what the Hawthorne effect is has come into dispute – in particular Mayo’s interpretation. The foundation of the dispute is beyond the scope of this thesis, however an interested reader is pointed to Draper (2006).

One problem with direct observation is that affect does not necessarily manifest itself through overt changes in behaviour, body language, or facial expression. This means that the observer will not necessarily be able to identify when a user is experiencing an affect. This is particularly true while the user is deeply engaged. For example, a computer game player may not show any observable changes while they are engrossed in a game.

Another issue with direct observation is being able to interpret observed changes as a specific affect. At times, it is apparent what affect the user is experiencing based upon their body language, behaviour, or facial expression. Two ‘easy to interpret’ observations are shown in Figure 7.5 and Figure 7.6.
However, the observation and classification of behaviour, body language, and facial expression is also subject to a tremendous amount of variability, uncertainty, conflicting signals, and evaluator bias. For example, when viewing Figure 7.7, one may be lead to believe that the girl is excited. However, it is also possible that the girl is frustrated, or perhaps simply stretching. Admittedly, the static nature of the image loses the context in which the ‘pose’ took place, but the point is that the subjective interpretation of body language, behaviour and facial expression is subject to personal interpretation.

The final point to make regarding direct observation of affective responses is that cultural variation does exist in the overt changes of behaviour, body language, and facial expression for a specific affective experience. This was an issue that Desmet (2002) experienced with the development of his Product-Emotion evaluation tool (discussed further in section 7.4.1). As a result, Desmet noted that there was a need to develop the tool in the context of each independent culture in order to address the cultural variation in the interpretation of body language. For direct observation and affect, it means that the evaluator must know how a specific culture expresses affect.
This is not a reasonable expectation for the evaluator given the diversity of users for any particular system.

7.2.2. Post-Use Evaluation Methods

Post-use evaluation methods allow users the opportunity to provide their overall opinions and suggestions regarding a product that they have recently used. The two post-use evaluation methods considered are interviews, and questionnaires. While these methods are discussed independently here, post-use evaluations are often of a combined interview and questionnaire format to create a more robust post-use evaluation method.

a. Interviews

Interviews are used as a method of engaging in dialogue with the user to elicit opinions and additional information about perceptions of the system (Shadbolt and Burton, 1995). Interviews are of two primary styles – structured and unstructured. Structured interviews are more rigid, where the evaluator has a set of defined questions that they ask the user. Though a dialogue could ensue from these questions, generally the user does most of the speaking with the evaluator recording responses. Conversely, unstructured interviews have general questions or themes of discussion outlined and encourage exploration of user responses within the scope of these questions and themes. With regards to system evaluation, many interviews use a combination of structured and unstructured styles. For example, a structured format may be used initially, and when (if) an interesting comment or answer is made by the user (perhaps relating to affect) then the evaluator may pursue and explore comment in an unstructured open dialogue with the user.

Structured interviews ensure that the evaluation stays focused – users respond to and provide input into the main areas of interest as identified beforehand. Structured interviews are limited in the fact that they do not allow for further probing of comments made by the user – which is one reason why interviews are often have a combination of structured and unstructured questions. Unstructured interviews, by themselves, provide the opportunity for users to express a range of thoughts and concerns in regards to a system and evaluators use their discretion when to probe further into these thoughts and concerns. With unstructured interviews, the user may provide input that is beyond the
scope of the evaluation, and may identify issues previously not considered within the context of the system. Throughout an unstructured interview, it is the evaluator’s responsibility to ensure that the aims of the evaluation are met.

Affective quality attribute scenarios for both structured and unstructured interviews are shown in Figure 7.8 and Figure 7.9. It should be re-iterated that the interview session is often a combination of structured and unstructured formats.

Both styles of interviews have similar limitations when eliciting information about user affective experiences. Asking users to recall a specific affective experience is not something people are accustomed to, and people may be reluctant to discuss feelings and personal experiences – particularly to an unknown evaluator (this was also an issue in the think-aloud protocol, 7.2.1.a). It is also known that retrospective recall of thoughts, feelings, and affect are not always accurate (Omodei et al., 1998). Therefore, any post-use statement by a user regarding their affect may not reflect the true experience had during product interaction.
b. **Questionnaires**

The use of questionnaires is common practice in usability evaluation. Questionnaires, as with interviews, can be structured (closed questions), unstructured (open-ended questions), or a combination of both. A stereotypical usability questionnaire would require the participant to rank various characteristics of the system on a Likert-style scale, ask the user to describe their thoughts and opinions on certain aspects of the system, and provide the opportunity to express other areas of thought or concern through open-ended questions.

Good questionnaires require time to develop to ensure that the questions being asked are specific or general enough (depending if it is an open or closed question), and to ensure that the questions target the evaluation objectives. Questionnaires also require pre-testing to ensure that the questions are properly understood by the users who will be completing them. However, after the questionnaire is developed they are normally quick, easy, and inexpensive to administer. An affective quality attribute scenario for questionnaires is provided in Figure 7.10.

![Figure 7.10: Affective Quality Attribute Scenario for Questionnaires](image)

As most questionnaires use a combination of open and closed questions they can ensure that the key questions in relation to the evaluation objectives are answered, while providing the user the opportunity to express their thoughts and feelings in regards to the system. Participants may show a greater willingness to express their thoughts and feelings on a questionnaire compared with an interview because they are not telling them directly to another person. By the same token, accurately describing personal experiences and feelings is difficult, and the user may not be able to explain what they actually felt or experienced during system use.
Similar to interviews, questionnaires require participants to recollect their thoughts, feelings, and experiences and this recollection may not accurately reflect the true experiences of the user (Omodei et al., 1998). In fact, the accuracy of affect recollection is an issue for all usability and QIU post-use evaluation methods.

7.2.3. The Software Usability Measurement Inventory (SUMI)

The SUMI is considered an instantiation of a closed-question questionnaire. As such, it does suffer the same limitations as other questionnaires in the context of affect. Of primary importance is that post-use evaluation methods require users to recollect their affective experiences, and recollections of this nature are considered unreliable (Omodei et al., 1998). However, the SUMI does identify affect as one of six dimensions that measures, and therefore warrants a focused discussion.

The development of the SUMI began in 1990 by the Human Factors Research Group at the University of Cork. The participants involved in the initial developmental study “were genuine end users who were using the software to accomplish task goals within their organisations for their daily work” (Kirakowski, 1994). Follow-up studies recruited participants from commercial and industrial contexts. These studies resulted in a questionnaire that reflects five dimensions of usability, which can be combined to provide a single ‘global usability score’: The five unique dimensions are:

- Efficiency
- Affect
- Control
- Helpfulness
- Learnability

Affect, as used in the SUMI, is defined as “the user’s general emotional reaction to the software -- it may be glossed as Likeability”. This suggests that the “dimension of affect” is broader than “affect” as defined within this thesis. This definition would seem to relate better to the concept of emotion than affect (see section 3.1.3 for distinctions between emotion and affect).
Within the SUMI, ten (of fifty) questions relate to their concept of affect. Examples of these questions include:

- Q 27. Using this software is frustrating.
- Q 42. The software has a very attractive presentation.

These questions relate to emotion and aesthetics respectively. Both emotion and aesthetics have been found to relate affect, but are not affect themselves (see 3.1.3 for the relationship between emotion and affect, and 3.3.2.c for aesthetics and affect).

This thesis considers affect to be the short-term immediate feelings in response to some stimuli, whereas the SUMI views affect simply as the general emotional reaction or ‘likeability’ of a system. This concept of ‘likeability’ is more suited to the concept of satisfaction, as discussed in section 4.4, where satisfaction is considered the general feeling that occurs while striving to achieve or after achieving some goal. To this end, the SUMI is identified within ISO 9241 as a tool being able to measure satisfaction (even though it is not one of the five unique dimensions – presumably the dimension of ‘affect’ is considered equivalent to satisfaction).

To summarise, the SUMI does not measure affect as it has been defined within this thesis. Rather, “affect” as used within the SUMI relates better to overall user emotion or satisfaction. Therefore, the SUMI is not considered further as a tool that can validate affect.

7.3. Requirements for Affect Validation

The evaluation methods described in section 7.2 provide a cross section of major evaluation techniques commonly used for QIU validation. While this cross-section of methods was not considered exhaustive, they were able to highlight several issues that hindered the ability of current evaluation methods from being used in relation to the validation of affect-based requirements. The key issues were:

- Interruption of the user during system use may alter their affective experience.
- Affect validation should not rely solely on the judgement of the evaluator to interpret affective experiences.
• The recollection of affect is unreliable.

These issues give rise to the following requirements for a method to validate affect:

• The validation measure shall be indicative of affective experiences.
• The nature of the evaluation shall not alter the affective experience of a user. In particular, the user shall not be interrupted during an evaluation.
• The validation method shall be unbiased by personal judgements on the part of the evaluator.
• Validation shall occur while the product is used to ensure that the affect recorded is representative of the actual affect experienced during use (and not based on recollection).

Unfortunately, current QIU evaluation methods do not meet all of these requirements. Therefore, it was necessary to explore alternative methods that may validate affect. These alternative methods are discussed in section 7.4.

In addition to the requirements stated above, any affect validation method should be viable for industry uptake. To be viable, the method shall not be cost prohibitive, and shall not require expertise that is beyond the scope of a normal system evaluator. The aim of this thesis is to provide an understanding of affect, how it relates to system quality, and to provide designers with guidance on how to design for affect and methods to validate affect.

7.4. Viable Affect Validation Methods

Three methods appear to meet the affect validation requirements set in section 7.3. The first method, user self-tracking, has been used during systems evaluation to gauge the user’s subjective perceptions of a system throughout system use (Swartz, 2001, Desmet, 2002; section 7.4.1). Psycho-physiology, the second method considered, assumes that there is a relationship between physiological changes and psychological experiences, including affect (Wilson, 2001, Healey and Picard, 1998; section 7.4.2). The final method, adopted from the field of naturalistic decision-making research, is the cued-recall debrief which re-immerses the participant in the activity they just performed.
allowing them to discuss their actions and thoughts freely, without the possibility of altering their actions (Omodei, Wearing, and McLennan, 1998; section 7.4.3).

### 7.4.1. User Self-Tracking

With respect to affect, user self-tracking requires the user to follow and record their affect while interacting with a product. What the participant actually tracks with respect to affect can vary, and may include one or a combination of:

- Tracking of the affective valence (positive or negative)
- Tracking of the arousal (strength) of the affective experience
- Tracking / identification of a specific affective experience

For example, one technique may track valence and arousal, another may track arousal and identify specific affects, and yet another may track valence, arousal and identification. For any of these combinations users may be required to track their affect continuously or at set intervals (e.g. every thirty seconds).

Swartz (2001) discussed participant self-tracking as a method that he and his colleagues were exploring to gauge user affect at different times during product use. They required users to identify specific affective experiences at regular intervals during product interaction. To facilitate the choice of affective experience, users selected from a set of facial avatars that best represented their current affect (e.g., a user may select an avatar representing ‘frustration’). The avatars provided a visual representation of an affective experience, and were used because of the recognised difficulties that users have when verbalising affective experiences. After this avatar is chosen it is placed on a timeline. This affect-timeline allows evaluators to track user affective experiences throughout interaction, and may allow them to identify specific product characteristics that contributed to a particular affective experience.

Desmet (2002) used a similar approach to gauge the experience users of domestic products had. Desmet’s approach differed in two manners. First, Desmet used ‘enhanced avatars’ – computer animated cartoons which began by expressing a neutral state and progressed to show an affective experience (e.g. starts neutral, progresses to joy). Desmet argues that people are able to better able to understand what an avatar
146

represents if they can see the progression to the affective experience, rather than a static representation of the affective experience. The second difference is that Desmet focused on relatively simple domestic products (e.g. a variety of kettles and other domestic appliances), and gauged their experience following exposure to each product. Desmet would show a user a product, let them interact with it, and following this, the user would then select from the series of animated avatars provided (and then continue with the next product). Desmet validated this method (labelled Product Emotion Measurement tool) in the context of simple domestic products, and found that it aided in eliciting experiences and discrete feelings (i.e. affect) related to the use of the product. The use of simple domestic products means that the extent of interaction was brief, and therefore a single measure post-use is sufficient to represent the affect experienced during the interaction. However, a single rating post-use is not considered sufficient to represent affect throughout an extended interaction – as normally occurs with most interactive systems. A single post-use rating could represent the overall experience, but would not represent the individual affects experienced during interaction (recall 3.2 – Why Affect and Not Emotion?). To use the Product Emotion Measurement tool it would be necessary to have users choose an avatar at regular intervals throughout interaction.

The methods used by Swartz (2001) and Desmet (2002) each required the user to identify a specific affective experience. As previously stated, it is also possible to have a user simply rate the valence and/or arousal level with respect to the affective experience on a scale. An example of such a scale is shown in Figure 7.11. This scale does not differentiate between specific affective experiences, but does provide a broad overview of user affect throughout interaction.

The primary issue relating to user self-tracking is that it will interrupt the user as it requires the user to perform an additional task, and this may alter user affect. Despite not meeting this requirement, it was still considered because of past-use by Swartz.
(2001) and Desmet (2002) as a method to evaluate affect. It can only be assumed that people will be able to track or identify their own affect. Participant self-tracking does not require the evaluator to provide any input, and therefore removes the possibility of bias on part of the evaluator. Self-tracking occurs during product interaction, and therefore satisfies the requirement of measures collected during use. Desmet’s Product Emotion Measurement tool is proprietary, and the cost of use may limit the viability. However, the method described by Swartz (2001) could be used without issues related to cost and availability.

Table 7.3 provides a summary of the affective validation requirements as met by user self-tracking.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure indicative of affective experiences</td>
<td>?</td>
</tr>
<tr>
<td>Does not interrupt the user</td>
<td>✗</td>
</tr>
<tr>
<td>Unbiased by the evaluator</td>
<td>✓</td>
</tr>
<tr>
<td>Collected during use</td>
<td>✓</td>
</tr>
<tr>
<td>Viable for industry uptake</td>
<td>✗/✓</td>
</tr>
</tbody>
</table>

The theorised quality attribute scenario for self-tracking is shown in Figure 7.12. To confirm the whether this method is suitable for affect validation a pilot study was conducted and is reported on in section 8.2. The aim of this pilot study was to determine if self-tracking caused a significant change in user affect, and if so then the method could not be used for affect validation.

![Figure 7.12: Affective Quality Attribute Scenario for Participant Self-Tracking](image-url)
7.4.2. **Psycho-physiology**

Psycho-physiology uses physiological signals as an indicator of psychological processes and responses. The most commonly known psycho-physiological tool is the polygraph. This tool can detect very subtle changes to a person’s physiological signals when that person is lying. Physiological signals that are frequently measured and used as an indication of psychological processes include galvanic skin response, blood pressure, and heart rate. These measures are non-intrusive, easy to collect, and the instruments required are relatively inexpensive. Methods using brain-wave patterns, obtained via an electro-encephalogram (EEG), are becoming more prominent but are not considered viable for general use as an affect validation method because of the costs associated with the equipment and the expertise required to interpret collected data. Eye-tracking is becoming more popular as a usability technique, and this may lend itself to tracking affect through pupillary response. However, current eye-tracking software is still limited, and pupillary response measurement is not common in these tools. Future versions of eye-tracking software may include this.

The link between affect and physiology is something that is accepted in everyday life. It manifests itself in everyday language used. For example, people may say “my blood was boiling because I was so angry”, or “she was blushing with embarrassment”, or “he was scared white”. References to the link between physiology, emotion, and affect can be found in literature dating as far back as two thousand years, where Erasistratos is recorded to have monitored a patient’s pulse concluded that he was suffering ‘unrequited love for his step-mother’ (as reported in Mesulam and Perry, 1972).

In modern day psycho-physiology research, the link between affect, emotion, and physiological change is widely accepted, though the exact process by which this link exists is subject to debate (see Black, 1970, Simonov, 1986, or Ball and Breese, 1999). Three of the most prominent theories describing the link between affect, emotion and physiology are the James-Lange theory, the Cannon-Bard theory, and Schacter’s Cognitive theory. These theories are summarised in Table 7.4. For more detailed descriptions, the reader is referred to the key references provided.
Table 7.4: Theories relating Affect, Emotion, and Physiology

<table>
<thead>
<tr>
<th>Theory</th>
<th>Key Reference(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>James-Lange</td>
<td>Rosenzweig and Leiman, 1982</td>
<td>Affect and emotions develop from the perception of physiological changes due to an external stimulus. That is, a physiological change occurs due to an external stimulus, and it is the sensation of this change that creates an affective experience, which is then cognitively processed to produce an emotional experience.</td>
</tr>
<tr>
<td>Cannon-Bard</td>
<td>Cannon, 1929, as cited in Coles, 1983</td>
<td>A stimulus is perceived and then cognitively processed resulting in an emotion appropriate for the context. The stimulus will also create a physiological response; however, this response does not influence the emotion.</td>
</tr>
<tr>
<td>Schacter’s Cognitive</td>
<td>Schacter and Singer, 1962</td>
<td>A stimulus is perceived, which creates an initial experience and also a physiological response. These are both cognitively processed resulting in an emotion appropriate to the context.</td>
</tr>
</tbody>
</table>

Each of these theories can be considered variations on the same theme. The original theory was the James-Lange theory, but this drew criticism from Cannon because the theory did not recognise the influence of context on the formation of emotion. Cannon then offered his own theory that recognised the influence of context but did not recognise the influence of physiological change on the formation of emotion. At the same time, Bard was forming his own theory which was found to be similar to Cannon’s (thus the label of “Cannon-Bard”). Schacter criticised the Cannon-Bard theory for not recognising the influence of physiological changes, and provided his own theory that can be considered a combination of both the James-Lange and Cannon-Bard theories.

The James-Lange and Cannon-Bard theories have minimal empirical evidence supporting them. However, Schacter and Singer (1962) performed empirical studies to validate Schacter’s Cognitive theory. These studies involved injecting participants with epinephrine – a naturally occurring hormone and stimulant within the body. The participants were divided into four different groups – one group not receiving the injection (control) and the remaining groups each given a different description about the side-effects of the epinephrine injection. One group was told the true side-effects (epinephrine-informed group), one group was told false side-effects (epinephrine-misinformed group), and one group was told that there would be no side-effects.
(epinephrine-ignorant group). They then proceeded to place participants into contrived social situations. In these situations, the epinephrine-informed group reacted the same as the control group. The epinephrine-misinformed and epinephrine-ignorant groups, on the other hand, reacted in an unexpected manner. Schacter and Singer concluded that the unexpected reactions in the misinformed and ignorant groups were due to the participants’ inability to understand why they were physiologically reacting the way they were, and therefore they attributed this reaction to the situation and altered their response accordingly. Conversely, the epinephrine-informed group understood that their physiological reactions were consistent with the effects of epinephrine, and therefore reacted normally. The outcome of this study showed that people do cognitively process physiological changes in the formation of emotion.

Figure 7.13 provides a model that shows the relationship between a stimulus, affect, and emotion. This figure integrates Schacter’s Cognitive theory and what is known about affect and emotion (as discussed in section 3.1). The primary difference between this and Schacter’s Cognitive theory is the cognitive processing of other influencing factors, as emotion is known to be influenced by factors beyond the sensation of a stimulus and physiological changes (refer to 3.1.1 for a description of other influencing factors).

![Figure 7.13: Stimulus through to Affect and Emotion](image)

Each of the theories described are ‘arousal based’, meaning that they assume that a physiological change corresponds to an increase in a specific physiological parameter (e.g. increased heart rate, increased galvanic skin response). However, this is not necessarily true. Coles (1983) provides descriptions of several studies where the physiological parameter actually decreases when an affective or emotional experience occurs. Therefore, a psycho-physiological measure for affect should accept increases and decreases in a physiological parameter as a possible indication of an affective experience.
Within the context of HCI, psycho-physiology has been explored by Wilson (2001), Healey and Picard (1998), and Ward and Marsden (2003). Wilson (2001) used physiological measures to identify the overall stress experienced when using products with varying media quality. Wilson manipulated video frame rate and audio degradations in animated media presented to a participant, and at the same time she measured the participants heart rate, galvanic skin response, and blood volume pulse (similar to blood pressure). She concluded that physiological signals differed depending upon the extent to which the degradations existed – the greater the degradation the greater the change in the physiological signal. Wilson’s analysis of data viewed each session as a single data source, so rather than tracking affect throughout system use it is more accurate to say that Wilson was using physiological measures as an indication of user emotion.

Using a different approach, Healey and Picard (1998) recruited an actress to explore the relationship between physiology and affect. The actress simulated eight distinct affective states, which included neutral, anger, hate, grief, love, romantic love, joy, and reverence. Healey and Picard acknowledge that these affective experiences are not generally the ones associated with product interaction. They justified these affects because they represent a range valence and arousal levels. Four physiological measures were taken while the actress was simulating these affective states – EMG (muscle activation levels), blood pressure, respiration rate, and skin conductance. They found that these physiological measures were unable to distinguish between specific affects, and it was not possible to discern positive and negative affect. However, they were able to discern between anger and ‘peaceful’ states and high versus low arousal states. One possible confounding factor is that Healey and Picard grouped ‘neutral’ under the category of peaceful, when this may have been better used as the control condition. Further, Healey and Picard measured the physiology corresponding to each affect independently. As such, it would be interesting to see if there were physiological changes while the actress was changing from one affect to another.

A study by Ward and Marsden (2003) used physiological measures to see if they could identify events that occurred during web-based interaction. They tracked a participant’s blood pressure, heart rate, and skin conductivity and sought to relate change in physiological patterns to significant events that occurred while using a web-based
system. The significant events that they identified were specific interrupting events, such as a pop-up box interrupting the performance of the participant. They noted that physiological patterns can be used as an indicator of these significant events, as heart rate and skin conductivity showed a marked increase when an event occurred. It was acknowledged by Ward and Marsden that this effect may be indicative of the flight-or-fight response, and may not relate to affect or emotion. Ward and Marsden (2003) also looked at the summation of physiological patterns over time and found that less-useful sites tended to result in an altered level of physiological activation – specifically a less usable tended to produce higher heart rates, higher skin conductivity, and lower blood pressure. These results are similar to those found by Wilson (2001).

Psycho-physiology appears to meet all the requirements for affect validation. The studies by Healey and Picard (1998) show that physiological measures can distinguish between differing levels of arousal, although they were unable to distinguish between different affects. The sensors that must be placed on the body to record physiological signals are non-intrusive and should not obstruct the user. Physiological measures are objective and unbiased, and are collected during use. Some physiology monitoring tools would not be viable for industry (e.g. electro-encephalogram), and others – such as eye-tracking software capable of tracking pupillary response – are not common. However, basic physiological monitoring tools (e.g. heart monitor, galvanic skin response monitor) are relatively inexpensive and there is software that assists in analysing recorded physiological data, which means that it is potentially viable for industry uptake.

Table 7.5 summarises the degree to which psycho-physiology meets the requirements for affective validation.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure indicative of affective experiences</td>
<td>✓</td>
</tr>
<tr>
<td>Does not interrupt the user</td>
<td>✓</td>
</tr>
<tr>
<td>Unbiased by the evaluator</td>
<td>✓</td>
</tr>
<tr>
<td>Collected during use</td>
<td>✓</td>
</tr>
<tr>
<td>Viable for industry uptake</td>
<td>✓</td>
</tr>
</tbody>
</table>
Previous works suggest that psycho-physiology could be a useful method to validate affect. It remains to be seen whether physiological measures will be able to discern changing affective states during use. The theorised quality attribute scenario for psycho-physiology is shown in Figure 7.14. The study conducted and described in section 8.3 provides a preliminary empirical evaluation regarding the ability of psycho-physiology as a validator of affect.

![Figure 7.14: Affective Quality Attribute Scenario for Psycho-Physiology](image)

**7.4.3. Cued-Recall Debrief**

The cued-recall debrief methodology has been adopted from the area of Naturalistic Decision making. Omodei et al. (1998) required a method that can elicit thoughts, affect, emotions, and cognitive experiences without interfering with a person’s behaviour in order to study the decisions and strategies of senior fire-fighters dealing with large bush-fires in rural Australia. Omodei et al. could not interfere with these fire-fighters as they were working in a time-critical situation. To this end, Omodei et al. (1998) developed the cued-recall debrief methodology.

This method involves a person, in their case a senior fire-fighter, wearing a small head mounted camera. This camera serves to record the visual field and audio environment of the participant. Immediately upon the completion of an activity session (or shift in the case of the fire-fighters) the video recording is played back to the person providing an own-point-of-view perspective. This own-point-of-view perspective is integral to the method, as it is an exact replication of how the person saw the situation previously. Omodei et al. believe that this perspective and the immediacy of playback would create an environment which encourages the person to become ‘psychologically re-immersed’ into the situation being played back. Theoretically, when re-immersed the person’s
thoughts, decision making strategies, emotions, and affect are recreated just as they were during the original activity.

Exploration and use of the cued-recall debrief methodology has revealed that people do become re-immersed in their activities (Omodei et al., 2002). While re-immersed the person can freely comment on thoughts, decisions, affect, or emotion without interrupting or changing their behaviour as their behaviour has already occurred. This re-immersion session is recorded to provide the evaluator with a detailed account of how the person perceived, interpreted, and reacted to the situation. Using this method, Omodei et al. (2002) have been able to gain a greater understanding of decision making processes by senior fire-fighters. The results that they obtain through the cued-recall debrief in the field setting is being used to aid in the development of a system that can aid in training fire-fighters in effective decision making strategies for bush-fire scenarios (Bushfire CRC, 2005). Research performed by Omodei and colleagues is a combination of computer-based simulations using specially designed software and actual field studies during real bush-fires.

The data collected through the cued-recall debrief methodology is considered ‘psychologically rich’ as it contains a vast amount of data regarding psychological processes of an individual during an activity or situation (Omodei et al., in press). The data collected are verbal reports that reflect a person’s thoughts, decisions, affect, and emotions. These reports can be quantified (e.g. grouping words into categories and comparing number of responses in each category), or qualitatively analysed (e.g. the evaluator reads the report and interprets the information contained). Omodei et al. (in press) encourage the qualitative analysis of this data as grouping and quantifying the verbal reports causes a loss of information and therefore does not provide the same degree of insight into the decision making strategies of an individual. However, as affect validation does not necessarily involve understanding overall thought processes or decision rationale, it may be appropriate to use quantification of verbal reports.

The cued-recall debrief methodology can be considered a hybrid of techniques described previously – it can be considered a combination of the think-aloud evaluation method (7.2.1.a) and post-use evaluation methods (7.2.2). It allows the user to provide the commentary normally elicited through the think-aloud method, but does not
interrupt behaviour as it occurs post-use. Using cued-recall debrief in the context of affect validation would require verbal reports related to affect (e.g. “I was getting excited at this point”) to be time-matched with what the user was doing. This would allow the evaluator to identify interaction characteristics to be identified and related to a specific experience.

The cued-recall debrief methodology meets most requirements for affect validation. There is always potential evaluator bias when analysing verbal reports. However, by focusing on affect potential bias is minimised because the evaluator’s interpretation is clearly based on user comments. For example, if a user says, “I was getting frustrated”, the evaluator should not interpret this further, and would accept the affect as frustration. To study decision making strategies (as Omodei et al., 2002 do) significant expertise on part of the evaluator is required to minimise bias.

The cued-recall debrief does not collect data during use. However, this requirement was formed on the basis that recollection of past affective experiences can be inaccurate (recall section 7.3). With cued-recall debrief the user is re-immersed into the activity, and as such the affective experiences are actually re-occurring in the user and are not recollections. To this end, the requirement should be changed to “Validation shall not rely on user recollections of affect”.

One possible issue with respect to industry uptake of the cued-recall debrief method is the duration of the evaluation session. Effectively, the time required for a single user to go through the evaluation process is double as the user must first interact with the product, and then go through the cued-recall debrief where they watch and comment on the entire interactive session. This extended duration may be restrictive in some contexts, and thus the cued-recall debrief methodology may not always be suitable.

Table 7.6 summarises the degree to which the cued-recall debrief methodology meets the requirements for affect validation as set in section 7.3.
### Table 7.6: Cued-Recall Debrief and Validation Requirements

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure indicative of affective experiences</td>
<td>✓</td>
</tr>
<tr>
<td>Does not interrupt the user</td>
<td>✓</td>
</tr>
<tr>
<td>Unbiased by the evaluator</td>
<td>x/✓</td>
</tr>
<tr>
<td>Collected during use</td>
<td>x</td>
</tr>
<tr>
<td>Viable for industry uptake</td>
<td>✓</td>
</tr>
</tbody>
</table>

Studies by Omodei et al. (1998 and 2002) have shown that the cued-recall debrief can collect information about a person’s affective experiences and is therefore considered a possible affect validation method. The theorised quality attribute scenario is shown in Figure 7.15. The study reported on in section 8.3 (in conjunction with the psychophysiology study) provides preliminary empirical results regarding the suitability of cued-recall debrief to validated affect.

![Figure 7.15: Affective Quality Attribute Scenario for Cued-Recall Debrief](image-url)
7.5. Summary

If affect is to be addressed within the context of system quality then it must possible to validate affect as part of QIU. External validation, ensuring that design characteristics known to promote user affect are integrated into the system, does not guarantee that the user will have or has had the desired affective experience. Only through QIU validation can it be ascertained whether a user has desired or anticipated affective experiences during system interaction.

This chapter emphasised the need to validate, and provided ideas regarding what an affect validation method must entail. In particular, this chapter:

- Discussed why affect needs to be validated within the context of QIU, and highlighted some issues that must be considered with respect to affect validation.
- Considered the entire process of validating requirements that relate to affect, using quality attribute scenarios as a guide.
- Critically analysed several QIU evaluation methods, focussing on their respective abilities with respect to the validation of affect. The methods considered were the think aloud methodology, direct observations, interviews, and questionnaires.
  - There was a focused discussion on the SUMI questionnaire as it identifies “affect” as one of its measures. It was found that the concept of “affect” within the SUMI relates to “satisfaction” or “emotion” in the context of this thesis.
- Provided a set of requirements that must be met for an affect validation method. These were:
  - The validation measure shall be indicative of affective experiences.
  - The user shall not be interrupted during an evaluation.
  - The validation method shall be unbiased by the judgements of the evaluator.
  - Validation shall occur during use.
During the cued-recall debrief discussion it was noted that this requirement should be changed to "Validation shall not rely on user recollections of affect."

- Proposed three methods that could be used to validate affect, namely:
  - Participant self-tracking
  - Psycho-physiology
  - Cued-recall debrief methodology

Chapter 8 continues on the theme of affect validation and details one pilot and one major study that evaluates the viability of the three proposed affect validation methods.
Chapter 8  Evaluating User Affect

This thesis first introduced system quality as fitness for purpose (see section 2.1). A quality product was said to be one that fulfils the user’s needs or wants (i.e. purpose) to a degree that the user perceives to be acceptable (i.e. fitness). If the user’s need or want is to experience a particular affect, or if the developer’s aim is for the user to experience a particular affect, then this becomes a requirement of the system. And as a requirement, it is necessary to validate it.

Chapter 4 through Chapter 6 explored ideas related to affect, design, and quality in use. Within these chapters, details were provided regarding design characteristics that could enhance the likelihood that the user will have an affective experience. However, designing a product that aims to create an affective experience in a user does not necessarily mean that the requirement was met. These chapters did not suggest how affect-related requirements could be validated.

Chapter 7 began to consider the validation of affective requirements. Using quality attribute scenarios as a framework several QIU validation methods were critically reviewed with respect to their ability to validate that a system could create user affect. From this review, a set of affect validation requirements were developed (section 7.3). Based on these requirements three methods – self-tracking, psycho-physiology, and cued-recall debrief – were proposed as possible methods to validate affect.

This chapter reports on two studies conducted. These studies assess the three proposed methods to evaluate user affect and validate affect-based requirements. Section 8.1 discusses what affect validation needs to achieve. A pilot study into the feasibility of participant self-tracking as a method to monitor and evaluate a user’s affective experiences is presented in section 8.2. A second more comprehensive study exploring the use of physiological measures and the cued-recall debrief as methods to evaluate a user’s affective experiences is then reported in 8.3.
8.1. Aims for the Evaluation of Affect

As discussed in the previous chapter, the evaluation of affect for purposes of requirement validation is difficult. It is difficult to observe – particularly in an objective manner – whether people are experiencing or have experienced affect. The critical review of current QIU validation methods provided a set of affect validation requirements (listed in section 7.3).

An ideal affect evaluation method would objectively identify the specific affective experience of a user (e.g. frustration, joy) and when this is occurring. This would allow the specific affect to be linked with certain characteristics or behaviour of the product, and is one of the reasons affect was chosen as the focus of this thesis rather than emotion (section 3.2).

However, objective measures do not appear capable of identifying the specific affective experience (i.e. they cannot tell the difference between frustration and joy?), but can provide an indication of when a person is experiencing affect. On the other hand, subjective measures – such as having a user describe their affect – can identify specific affective experiences and when they are occurring, but rely on the user to identify these. For subjective measures, it is not possible to guarantee that the user will describe all affective experiences, and may even choose not to describe an affect.

The three methods considered for affect validation (i.e. participant self-tracking, psycho-physiology, and cued-recall debrief) are summarised in Table 7.3 with respect to the ability to identify or indicate affective experiences.

| Table 8.1: Affect Validation Methods and Identification or Indication of Affect |
|---------------------------------|-----------------|-----------------|
| **Method**                      | Identification of Specific Affective Experiences | Indication of Affective Experience |
| Participant Self-Tracking       | Subjective      | Subjective      |
| Psycho-Physiology               | Subjective      | Objective       |
| Cued-Recall Debrief             | Subjective      | Subjective      |

Participant self-tracking and cued-recall debrief do not allow for objective measures to be taken – they rely on the subjective responses of the user to identify specific affects and when they occur. Psycho-physiology allow for the objective indication of when
user affect is occurring, but cannot differentiate between affects (as demonstrated by Healey and Picard, 1998).

8.2. Pilot Study i User Self-Tracking

Section 7.4.1 described two participant self-tracking methods previously used to track affect. Desmet (2002) used animated avatars to assess user experiences at the end of a short-term interaction with a variety of domestic products (e.g. kettle). Swartz (2001) required users to select a facial avatar that represented their current affective experience at regular intervals during system use, and this avatar was then placed on a timeline. As participant self-tracking was used to track and record user affect within a system-use context, it was considered a possible affective requirements validation method.

A pilot study was performed to explore the viability of several user self-tracking options while watching a comedy video. These options were:

- **Option 1:** The participant selecting from a set of 24 facial avatars at regular intervals (every 30 seconds, prompted by the experimenter if the participant missed). These avatars are then placed on a timeline. (This mimics the method used by Swartz, 2001).
- **Option 2:** The participant selected from a set of 24 facial avatars whenever they experienced a change in their affective experience. This avatar was then placed on a timeline.
- **Option 3:** The participant continuously tracked the degree of valence of their current affective experience on a scale labelled positive – neutral – negative. This continuous tracking was recorded by video (to allow for later transcription of affective experience change). This is similar in nature to the physical knob / slider often used to assess the reaction of a TV audience.
- **Option 4:** The participant added or wrote verbal labels describing their affect on a timeline at regular intervals (every 30 seconds, prompted by the experimenter if the participant missed).

The video watched was a segment from the series titled “Yes! Minister”. A video was used to allow all participants to complete the self-tracking exercise in unison. A comedy video was chosen as it was anticipated that a comedy would create the greatest
number of positive affective experiences in a user, and the nature of this comedy (i.e. British) was suited to the participant demographic. Further, the content of the video related to the design of surveys and questionnaires, and was a topical video for the HCI discussion group. It was anticipated that the nature of comedy and the topical content would increase the engagement of participants, and subsequently would increase the likelihood of affective experiences.

The aim of the pilot study was to identify which, if any, of the self-tracking options can effectively track user affect.

8.2.1. Method

a. Participants

The pilot study involved eight participants, with two participants independently using each self-tracking option described. The participants were attendees at a weekly HCI research discussion group at an Australian University. The participants included academics, students, and general staff.

b. Procedure

The pilot study was conducted in three stages. The first stage informed the participants of the pilot study aims (i.e. which method can be used to track user affect), and briefly described the issues related to the validation of affective requirements. During this stage, all four self-tracking methods were described, and any queries were answered.

The second stage required all participants to self-track their affective experiences while watching a ten minute segment from a comedy video.

The concluding stage was a focus group session. The aim of this session was to allow participants to freely comment on their perceptions regarding the viability of the self-tracking methods for use in the evaluation of user affect.

c. Equipment

The study took place in a standard meeting room, with the "Yes! Minister" video played on a standard VCR and television. Four self-tracking devices were used in this study, with two participants using each. Appendix D: provides samples of each device.
### 8.2.2. Results

Table 8.2 summarises observations of the experimenter and key comments raised by participants during the focus group for each self-tracking option.

**Table 8.2: Summary of Self-Tracking Observations and Participant Comments**

<table>
<thead>
<tr>
<th>Self-tracking Option</th>
<th>Comment or Observation</th>
</tr>
</thead>
</table>
| Avatars, 30 second intervals            | • Observation: The 30 second intervals were too short to select an avatar – by the time an avatar was chosen, it was almost already time to select again. A reduced set of key avatars may be beneficial if this method is to be used.  
  • Comment: The regular interval of 30 seconds disrupted engagement with the video. Participants felt that this altered the affective experiences reported.  
  • Comment: The avatars provided a reference to the participant regarding the range of affective experiences they could have. |
| Avatars, self-regulated intervals        | • Observation: There was a tendency to forget to update affective experiences through self regulation  
  • Comment: The avatars provided a reference to the participant regarding the range of affective experiences they could have. |
| Continuous tracking of valence           | • Comment: The concentration required to continually track affective experiences means it is not possible to become fully engaged in the video  
  • Comment: If the participant chose to concentrate more on the video than the self-tracking, then the participant forgot to continually update their affective experience valence |
| Verbal labels, 30 second intervals       | • Observation: The regular interval of 30 seconds disrupted engagement with the video. Participants felt that this altered the affective experiences reported.  
  • Comment: It was difficult for the participant to label their current affective experience. Participants reported looking at the avatars (provided to other participants) for ideas on how to describe their current affect. |

### 8.2.3. Discussion and Outcomes of the Pilot Study

All options explored for affective experience self-tracking appear to be inadequate for use as an evaluation method for user affective experiences. In this study, participants were only required to watch a video and record their affect, and each experienced
problems doing so. By comparison, using an interacting product requires the user to perform additional tasks, which would serve to exacerbate the issues noted in this study.

The options that required the participant to respond at regular intervals disrupted the engagement with the video, and participants felt that this disruption adversely impacted upon their affective experience. This was previously recognised and integrated into the affect validation requirements identified in section 7.3; and it was also known that self-tracking was deficient with respect to this requirement (see Table 7.3). However, as self-tracking with regular intervals was previously used to track affect (e.g. Swartz, 2001) it was important to test this method. The results from this study indicate that the additional task of self-tracking has a large degree of influence on affective experiences.

Participants who performed the self-regulated interval or continuous tracking options found that they neglected updating their affective experiences if they chose to concentrate on the primary task – watching the video. This has clear consequences when trying to track, record, and evaluate user affective experiences.

Although expected (based on previous studies performed by other researchers), one positive outcome of this study was the finding that the avatars did assist users in identifying or labelling their affective experiences. Therefore, if there is a need for the user to identify specific affective experiences it may be useful to provide them with avatars showing a variety of different affects. However, within the study, it was found that by the time the user chose an avatar it was often time to select again (and thus the participant did not have a chance to engage in the video). Therefore, a set of avatars should be provided, but these avatars should only represent key affects to minimise time. A sample set of avatars representing twelve affects (and one representing neutral affect) is provided in D.5.

8.2.4. Conclusion

This pilot study was intended to see if self-tracking of affect could be used as a method to evaluate user affect. It was intended to explore whether previously identified problems (e.g. interruption of the user) would manifest themselves during the evaluation of affective experiences, as these problems were not experienced in other similar studies. However, the results of this study suggest that these previously identified
problems and some newly identified problems do change the affect experienced. Therefore, self-tracking of affect does not appear to be capable of being used in the evaluation of user affect, and probably cannot be used as a QIU validation method for affective requirements.

8.3. Study i  

Psycho-physiology and Cued-Recall Debrief

Psycho-physiology results from this study were published in Bentley, Johnston, and von Baggo, 2003. Cued-recall debrief results from this study were published in Bentley, Johnston, and von Baggo, 2005. Complete results are provided here.

Psycho-physiology and the cued-recall debrief methods were discussed as possible methods to evaluate user affective experiences in section 7.4. Psycho-physiology can indicate when a user is experiencing affect, but cannot differentiate between affects. Through cued-recall the user reports their specific affects, however this is subjective and relies on the user to recognise and want to report these affects (i.e. if a user does not want to report a specific affect in fear of embarrassment). Further, neither psycho-physiology nor cued-recall have been used to assess user affect during product use.

As discussed in section 8.1, the ideal evaluation method would provide the objective identification of user affect. This was not considered possible using a single method. Cued-recall debrief can subjectively identify user affect, and psycho-physiology can objectively indicate when a user is experiencing affect. Combined, these two methods appear to be an optimised affect evaluation method. It allows for the objective indication of when an affective experience occurs, while subjectively providing insight into the exact nature of the affective experience.

As neither of these methods has previously been used to evaluate user affect during system use, the combined use provides the opportunity to cross-check their ability to identify or indicate affective experiences. For example, if the physiology indicates that the individual is having an affective experience and during the cued-recall debrief they identify an affective experience at that time it confirms that each method is able to identify or indicate affect. On the other hand, if one method is indicating an affective experience but the other is not then there may be some issues related to the reliability of the methods.
The study described here assesses psycho-physiology and the cued-recall debrief method with respect to their ability to evaluate user affective experiences. Within this study, participants were asked to play two computer games. Computer games were chosen as the product of interest because it is expected that these will create a number of affective experiences in a user, allowing greater opportunity to study the relationship between affect, physiology, and system characteristics (more detail regarding why computer games were chosen is in section 6.1). Physiological measures – heart rate, skin perfusion, and breath rate – were recorded while playing. These three measures were chosen because they are not intrusive to the participant, and do not require a high level of expertise to use. As such, each could be used in an industrial environment. After playing each game the user undergoes a cued-recall debrief. This allows associations to be made between user-identified affect (through the cued-recall) and physiological measures.

Section 8.3.1 formalises the purpose and aims of this study. The methods used are described in section 8.3.2 – a pilot study reported on in section 8.3.3 was conducted to ensure the methods used were optimal. Results from the main study are presented in sections 8.3.4 and 8.3.5 with a high-level summary of these results in 8.3.7. Section 8.3.8 provides a qualitative review of the physiological measures used and the cued-recall debrief method with respect to affect evaluation. Section 8.3.9 concludes this study, and identifies some areas of future research.

### 8.3.1. Purpose

While past studies seem to indicate that psycho-physiology and cued-recall debrief would be suitable evaluation methods for user affect, it was not known if these methods can be used to evaluate user affect during system use.

The primary aim of this study was to identify whether psycho-physiology and/or cued-recall debrief could be considered a viable evaluation method for user affective experiences within the context of system use. Specifically, this study will determine whether:
• Psycho-physiology can provide an indication of when an affective experience is occurring during system use; and
• Cued-recall debrief can be used to elicit information regarding specific user affective experiences during system use.

The outcome of this study will address whether these methods could be a viable QIU validation method for affective requirements. Results related to the primary aims are presented in section 8.3.5. It is expected that both methods can be used to evaluate affect, as each appear to meet all requirements set forth in section 7.3. It is believed that an affective experience identified during a cued-recall debrief will have physiological changes associated with it.

Several secondary aims were also identified. These secondary aims sought to determine the limitations of psycho-physiology as a method to evaluate affect should it prove useful. The secondary study aims to determine what, if any, effects the following variables have on user physiological characteristics in response to affect:

• Affective valence: Is it possible for physiological measures to distinguish between positive and negative affect?
• Type of system used (i.e. game type): Do the physiological characteristics differ based on the type of game played?
• System preferences (i.e. game preference): Do the physiological characteristics differ based on user preference for a certain game?

These secondary aims will provide greater understanding with respect to the relationship between user affect, physiology, and system use. Results related to the secondary aims are presented in section 8.3.6.

8.3.2. Method

Summary of Method: the study involved ten participants. Each participant was asked to play two different computer games for ten to twenty minutes each (one was a Tetris style game, the other was a car racing game). Computer games were chosen as the product to use for reasons described in section 6.1. While playing each computer game, the participant’s heart rate, skin perfusion, and respiration rate were monitored.
Following the game play session, the participants were then guided through a cued-recall debrief session. Debrief comments and physiological data were then related and analysed.

a. Participants

Ten participants were recruited through paper-based advertisements posted in buildings at an Australian University. These advertisements informed potential participants that they would receive $40 remuneration, and that the study would take approximately two hours. The only restriction was that all participants must be between the ages of eighteen and twenty-nine years of age. This restriction was in place to ensure a relatively homogeneous population to minimize the amount of physiological differences known to exist due to age differences.

Participants completed a pre-test questionnaire to obtain demographic data (this questionnaire is available in E.4). Along with general background questions, the demographic questionnaire contained questions related to the participant’s computing habits and computer game playing preferences. Key participant demographics are shown in Table 8.3.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Occupation</th>
<th>Computer Use (hours/week)</th>
<th>Leisure Computing (hours/week)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>Student</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>Student</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
<td>Musician</td>
<td>&gt;10</td>
<td>2-5</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>Education</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>5</td>
<td>Male</td>
<td>Chef</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>6</td>
<td>Female</td>
<td>Student</td>
<td>&gt;10</td>
<td>&lt;1</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>Student</td>
<td>&gt;10</td>
<td>&lt;1</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>IT industry</td>
<td>&gt;10</td>
<td>6-10</td>
</tr>
<tr>
<td>9</td>
<td>Female</td>
<td>Student</td>
<td>&gt;10</td>
<td>2-5</td>
</tr>
<tr>
<td>10</td>
<td>Male</td>
<td>Education</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
</tbody>
</table>

*Leisure computing was defined as the amount of time that participants used a computer for non-work purposes. Examples provided to the participant included surfing the internet, playing games, etc.*
None of these participants had experience using the computer games selected for use in this study, although nine participants had used a game similar to the Tetris-style game. Only two participants had experience with racing-style games. As such, all participants were considered novices with respect to the specific games used in this study.

b. Procedure

Each participant was given an information sheet to read when they first arrived (included in E.1). Key points within this information sheet were verbally described to the participant, and the participant was given the opportunity to ask any questions that they may have. The participant was then asked to sign an informed consent form acknowledging that they understood the study and their individual rights (see E.2).

Upon completion of the consent form, the physiological monitoring equipment was attached to the participant, as shown in Figure 8.1. Equipment used included a thermistor capable of monitoring breathing, a pulse oximeter to monitor heart rate, and a laser-Doppler perfusion machine capable of monitoring skin perfusion. Details of how physiology data was collected are described in section 8.3.2.c, and details of the equipment used are in section 8.3.2.e.
Participants were then asked to sit quietly for five minutes. During this time, baseline physiology measures were taken. Only the last two minutes of this data was used for the baseline measure. The initial three minutes allows the participant to become used to the equipment, and to minimise the influence of artificially raised physiology measures due to nervousness or apprehension. These baseline measures were subsequently used to normalise the physiological measures to a percentage of baseline values, allowing comparison between different participants with differing baseline values.

Following the period of quiet sitting, the participant was introduced to the first computer game that they were to use during the session – participants were randomly given a racing or Tetris-style game first (see section 8.3.2.e for game details). The researcher explained the aims and controls for the game to the participant, after which no further feedback or input was provided by the researcher.

The participant was asked to play the first game for ten to twenty minutes. The exact time reflected how long it took a participant to complete a game, as the researchers did not want to interrupt the participant mid-way through. During the play session, researchers did not make any comment or provide any feedback to participants. Physiological signals were continuously monitored to ensure that the equipment and sensors were continuing to track the physiological variables. The entire game playing session had video and audio recordings made – the set-up of the lab and recording equipment is described in section 8.3.2.e.

Following each game play session, the participant was immediately taken through a cued-recall debrief. The participants were situated in front of a television screen and watched themselves play the computer game from an own-point-of-view perspective (as required by the cued-recall debrief method – see section 7.4.3). The participant was encouraged to speak freely regarding all aspects of their interaction and was not censored in any way during the debriefing process. To help focus the discussion on the identification of affective experiences, the participants were provided with a set of avatars that represented twelve different affects (provided in D.5). This cued-recall debrief session was video-taped with audio. After the first debrief session, the participant was introduced to the second game. The procedure was then repeated, starting at the point where the participant received a tutorial on how to play the game.
At the completion of the overall session, the participant was given the opportunity to ask any questions in relation to the study performed. The participant was asked to complete a permission form giving the researchers the option to publish short clips or images from their video for use in publications related to this body of research (E.3). This added consent was due to the unique nature of this study, and would allow the researchers to publish images or video to aid in communication about the study set-up.

c. Physiology Data Collection Method

Participants had their heart rate, skin perfusion, and respiration flow monitored and recorded throughout the game-play session. All monitoring equipment was attached to an analogue to digital converter, enabling the analogue data collected to be output to a PC using appropriate data collecting software. All physiological data was recorded at 256 Hz (i.e. 256 data points each second). All measures were recorded from the right side of the body for consistency. A schematic of the physiology recording equipment is shown in Figure 8.2. Precise details regarding the physiology equipment is available in section 8.3.2.e.

![Figure 8.2: Physiology Equipment Schematic](image)

The heart rate was recorded using a pulse oximeter. This can provide a heart activation tracing, pulse rate, and blood oxygenation levels – for this study only the pulse rate was recorded. The pulse oximeter was attached to the participant’s right thumb. It is generally recommended to attach the pulse oximeter to the index or middle finger as the thumb creates some difficulty in accurately recording blood oxygenation. However,
blood oxygenation levels were not relevant to this study, and as participants were required to use their fingers (thumb excluded) it was considered appropriate to place the pulse oximeter on the thumb.

The skin perfusion was monitored using a laser-Doppler flow measurement device. This measurement provides the same information as Galvanic Skin Response, but is less sensitive to participant movement and provides greater accuracy of measurement. The laser diode was attached to the participant’s right forearm on the medial side. This location was chosen because it contains a large number of vascular vessels close to the surface of the skin, and which is optimal to monitor changes in skin perfusion.

The respiration flow was monitored using a nasal thermistor. The thermistor was placed just under the nose where it responds to the small ambient temperature changes when a person inhales or exhales and thus produces a resulting output of increasing and decreasing temperatures corresponding to the breathing pattern (exhale – warm air; inhale – cool air). This was converted to a breath rate by the Chart 4 software that identified consecutive maximum and minimum temperature points. The nasal thermistor is considered sensitive enough to pick up temperature changes from both mouth and nose breathing.

The Chart 4 software used to collect physiological data on the pc allowed for time-stamped comments to be made during data collection. This feature was used to mark the beginning and end of a game-play session. By marking this, it also provided an identifiable mark in the video to which the video time can be calibrated.

d. Data Analysis Procedure

To begin the data analysis, it was first necessary to transcribe and time-stamp each the cued-recall debrief comment. Each comment was then classified as one of:
• **Positive Affective Comment**: debriefing comments where the participant indicated that they had a positive feeling or affective experience at that time (e.g. "I was excited here");

• **Negative Affective Comment**: debriefing comments where the participant indicated that they had a negative feeling or affective experience at that time (e.g. "I was getting frustrated");

• **Neutral Affective Comment**: debriefing comments where the participant indicated that they were not having an affective experience at a specific time (e.g. "I wasn't really thinking too much here, I was pretty much just going along");

• **Generic Comment**: debriefing comments where the participant looked at the experience as a whole (e.g. "I already suck at driving, so I knew that I would already suck at computer driving");

Only the positive, negative and neutral affective comments were examined further. Non-affective comments were treated as the control condition – times when the participant was not having an affective experience. Positive and negative affective comments were combined for the primary analysis (which was to determine if psychophysiological measures could distinguish between neutral and affective experiences). The valence (positive or negative) was used in secondary analysis of the data that sought to determine the limitations of physiological measures as they relate to affect. The classification of comments was initially performed by the researcher. However, to ensure that this classification was correct three other independent researchers also classified the comments. These three researchers had psychology backgrounds, and were familiar with classifying subjective comments. The four different classifications were checked for reliability (see section 8.3.4 - Inter-Rater Reliability).

Following transcription and classification, the cued-recall debrief comments were combined with the physiological data. Initially, all physiological data was imported to an MS Excel Spreadsheet. MS Excel was used to assist in time-matching the cued-recall comments to the physiological data. Only physiological data collected during the period 25 seconds before and after each relevant comment was retained (50 seconds in total). This was necessary to reduce the amount of data to be processed. As the data
was recorded at 256 Hz for each individual comment there were still 12 800 data points for each physiological variable (50 seconds * 256 data points/second). A lower collection frequency could not be used, as this would reduce the fidelity of the data. The 50 seconds of physiological data was divided into five intervals as shown in Figure 8.3, with the third interval representing the time when the affective or non-affective comment was made. These intervals aid in describing effects at a certain time.

MS Excel was also used to process the baseline data for each participant. This baseline data was used to convert all physiological data for each participant into a percentage of baseline, allowing for comparison between participants.

After initial data manipulation, reduction, and conversion to percentages of baseline, all data was exported to SPSS v11.5 for statistical analysis. Pallant (2001) provided the statistical analysis procedures used for this study. At times, the procedures given by Pallant were extended to include additional analyses (particularly post-hoc analyses). A complete flow-chart of the analysis process is shown in Figure 8.4, and a description follows.
Within SPSS, the first step was to identify all outliers within the data set. Outliers were defined as any data point that was greater than two standard deviations from the mean. Outliers that were found to be three standard deviations from the mean were removed immediately as these likely represent electrical artefacts of the physiological measurement, and not the participant’s physiology. Outliers within three standard deviations were reviewed on a case by case basis, with the main judgement criterion being whether the data was continuous or whether the data was irregular (an example shown in Figure 8.5).
Irregular outliers shown (as circled) would be removed from the data set.

Following the removal of all outliers, mixed between-within subject analysis of variances (also known as split-plot ANOVAs or SPANOVAs) were performed. The ‘between’ comparison was whether the physiological data was able to differentiate between affective comment and non-affective comment conditions for each of the time intervals. The within analysis involved checking whether subsequent intervals showed a significant increase or decrease. The tests conducted within each SPANOVA are shown in Figure 8.6.

SPANOVAs were performed on each variable (heart rate, skin perfusion, breath rate). The SPANOVA effectively performs five ‘between-condition’ analyses (affective vs. non-affective) and four within condition analyses for each comment, as shown in Figure
8.6. For each variable, a SPANOVA was performed for averages, variance, and absolute change. These are described as:

- **Averages**: The actual values represent the 10 second average of the physiological variable as recorded.
- **Variance**: The variance values represent the variance of the physiological variable over a 10 second period.
- **Absolute Change**: The absolute change values are representative of the maximum change that occurred within the interval. It is calculated by taking the maximum value within the interval and subtracting the minimum. To ensure that the maximum and minimum values are not miscellaneous artefacts caused during data collection, it was required that the maximum and minimum values be consistent for at least one second in duration.

If an affective or non-affective comment was elicited within 25 seconds of another comment the data was removed from initial analysis. This was done to ensure that the initial analysis results were not influenced by adjacent affective experiences (i.e. to ensure that the physiological data was not already changed due to another affective experience). Follow-up analyses were conducted which focused only on physiological trends leading up to or following an affective experience. These follow-up analyses allowed for a greater number of comments to be included in the analysis, increasing the statistical power. However, the generalisations made based on these follow-up analyses are not as generalisable as the initial analysis because they only describe ‘half the picture’. This is shown in Figure 8.7.
In total, nine SPANOVs were performed for each cued-recall comment in the initial and follow-up analyses. Post-hoc tests were performed whenever a significant effect (main or interaction) was found within any SPANOVA. Post-hoc testing included one-way within subject ANOVAs or one-way between subject ANOVAs. If significant effects were found in the one-way within subjects ANOVA paired t-tests using Bonferroni adjustments were conducted to ascertain the exact nature of the effect.

e. Equipment

The equipment is divided into four sub-sections:

- Laboratory set-up;
- Physiological recording equipment;
- Game-play session equipment; and
- Cued-recall debrief equipment

Laboratory Set-Up

The study took place in the SCHIL Usability Laboratory at Swinburne University of Technology. This lab has been designed to be highly flexible and adaptable, allowing for a wide variety of projects and studies to be carried out for both academic and industrial purposes. The lab was set up as shown in Figure 8.8. It included two cameras, two microphones, two computers, and a ‘debrief station’, which was composed of a television and two video recorders.
Physiology Recording Equipment

Table 8.4 details the equipment used to collect physiological data.

<table>
<thead>
<tr>
<th>Item</th>
<th>Description of Device</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection PC</strong></td>
<td>This computer was used to collect data from the physiological monitoring equipment.</td>
</tr>
<tr>
<td>233 MHz Pentium PC, 128 mb RAM, 17 inch CRT</td>
<td></td>
</tr>
<tr>
<td><strong>A/D Converter</strong></td>
<td>This device allows the conversion of up to eight analogue signals to digital format for collection by a PC. Four signals were collected: heart rate, heart activation tracing, tissue perfusion, respiratory flow.</td>
</tr>
<tr>
<td>AD Instruments Powerlab 8/s</td>
<td></td>
</tr>
<tr>
<td><strong>Data Collection Software</strong></td>
<td>This software enables a computer to interpret, store, and visually present data collected through an A/D converter. It also allows comments to be added during or after data collection. Data can be exported to various formats (including MS Excel), and contains some signal processing algorithms.</td>
</tr>
<tr>
<td>Chart 4, version 4.1.2 by AD Instruments</td>
<td></td>
</tr>
<tr>
<td>Heart Rate Monitoring Device</td>
<td>Nonin 8600 Pulse oximeter</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>This pulse oximeter is able to monitor and output details of blood oxygenation levels, heart rate, and heart activation tracing.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Skin Perfusion Monitoring Device</th>
<th>Transonic Systems' BLF 21-Series</th>
</tr>
</thead>
<tbody>
<tr>
<td>The laser-Doppler skin perfusion monitoring device provides a measure of microcirculation in 1 mm³ of tissue. This value is representative of effects due to blood pressure and sweat responses.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Respiratory Flow Monitoring Device</th>
<th>Nasal Thermistor and Voltmeter (generic brand)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The nasal thermistor is a small probe that is sensitive to very small temperature changes. It can monitor respiration by identifying ambient temperature changes around the nose during inhalation and exhalation.</td>
<td></td>
</tr>
</tbody>
</table>

### Game Play Session Equipment

Two different styles of computer games were used – one action/racing oriented (Need for Speed 5 by EA Games, 2001), and one puzzle oriented (Tetris clone by MiniClip, 2001). None of the participants had experience with either of these games.

During the game play sessions video composed of the participant’s body, participant’s hands, the participant’s monitor (via screen capture device), and the evaluator’s monitor (via screen capture device) was recorded using a quad-display mixer. The resulting video has the characteristics shown in Table 8.5. All audio from this session, including in-game audio and participant comments or exclamations, were recorded on the left audio channel of this video (as described in Table 8.6).
Table 8.5: Video Recording Details, Game Playing Session

<table>
<thead>
<tr>
<th>Video Recording Details (Quad Display)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants Body (Camera 1, C1)</td>
</tr>
<tr>
<td>Participants Hands (Camera 2, C2)</td>
</tr>
<tr>
<td>Game Screen (Participants Monitor, Screen Capture)</td>
</tr>
<tr>
<td>Data Screen (Evaluators Monitor, Screen Capture)</td>
</tr>
</tbody>
</table>

Sample of Resulting Video

Table 8.6: Audio Recording Details, Game Playing Session

<table>
<thead>
<tr>
<th>Audio Recording Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right Channel Audio (A_r): Blank</td>
</tr>
<tr>
<td>Left Channel Audio (A_l): Game-play session audio</td>
</tr>
</tbody>
</table>

Cued-Recall Session Recording Equipment

The cued-recall debrief session re-recorded the video and audio from the game play session. The debrief audio (comments made by the participant) were added to the original video on the right audio channel. This required two video recorders to be used, with one playing the original video (with audio) and the second recording video and left channel audio input from the first video recorder, and microphone input from the cued-recall debrief session to the right channel of audio. The final video had the same visual properties as the original (Table 8.5), but with two channels of audio (see Table 8.7).

Table 8.7: Audio Recording Details, Post-Cued-recall Session

<table>
<thead>
<tr>
<th>Audio Recording Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right Channel Audio (A_r): Debriefing Audio</td>
</tr>
<tr>
<td>Left Channel Audio (A_l): In game Audio</td>
</tr>
</tbody>
</table>

The resulting single video provided audio from both the game-play session and the cued-recall debrief session in addition to complete video from the game-play session.
This greatly aids transcription of participant comments, as all comments are already time-matched with video and physiological data (as appearing on the video). Further, by muting the left channel of audio, it was possible to focus on the debrief comments when the speaker was difficult to hear.

8.3.3. Pilot Study

A pilot study was conducted prior to the official study to ensure that the procedure was optimal. Two participants were involved in this pilot study, and followed the same procedure as described in section 8.3.2.b. These participants were colleagues of the researchers, and did not receive any remuneration for their involvement. The pilot study participants were encouraged to talk candidly about the method and to describe aspects of the method that they felt could be improved.

The original methodology (pre-pilot test) did not include using the set of avatars presented in D.5. During the cued-recall debrief, the first pilot study participant experienced difficulties when verbalising her affective experiences. The researcher asked the participant if she would find a visual aid with several different affects pictured useful, to which she replied positively. The second participant was given this aid, and he referred to it regularly during throughout cued-recall debrief session. Based upon this, it was decided that this aid would be provided for all participants in the study. This was the only change made to the method.

The pilot study was also used as an opportunity to informally check whether physiological measures appear to change in response to affective experiences. There were positive indications that physiological measures do correspond to user affective experiences noted during the cued-recall debrief. Due to the limited sample size of two in the pilot study, there was no statistical significance found.

8.3.4. Inter-Rater Reliability

Prior to analysis of data, it was necessary to ensure that the classification of cued-recall debrief comments was reliable. If it is not possible to reliably classify affect-related comments then cued-recall cannot be used as a method to evaluated affect.
To check the reliability of affect comment classification, four researchers independently classified the comments as positive affect, negative affect, neutral affect, or generic – as described in section 8.3.2.d. The four researchers included the one conducting this study, and three other individuals who had a psychology background and familiarity with the classification of verbal reports. The form used to classify comments is located in E.6. The results from each researcher are provided in E.7, and the inter-rater reliability SPSS analysis is in E.8.

The four classifications showed a very high reliability result (Cronbach’s alpha = .9831). This indicates that the classification was highly reliably. Therefore, it is possible to state that the classification of cued-recall debrief comments can be done consistently. If this was not possible, then cued-recall could not be used as a method to evaluate affect (as without reliability it is impossible to have validity in a measure).

8.3.5. Results – Targeting Primary Study Aims

The primary aim of this study was to determine if psycho-physiology and/or cued-recall debrief can be used as a method to evaluate user affective experiences within the context of system use (recall 8.3.1 - Purpose). As neither of these methods have been used for this, it was necessary to ‘self-validate’ – that is, if one method showed a result then the other method must also as well. It is expected that an affective comment made during the debrief will be related to some change in user physiology, and that a neutral comment will be related to no change in user physiology. If these do not occur, then one or both of these methods cannot be used for this purpose.

The results in 8.3.5.a report on the initial analysis. The initial analysis focused only on comments that did not have another comment made in the 25 seconds before or after. Section 8.3.5.b reports on the follow-up analyses, and included:

- Physiological trends leading up to an affective comment if there was no other comment made in the 25 seconds prior to the comment of interest (regardless of whether or not there was a comment following)
- Physiological trends following an affective comment if there was no other comment made in the 25 seconds following the comment of interest (regardless of whether or not there was a previous comment).
These follow-up analyses were performed to increase the statistical power of the debrief comment-physiological response relationship, and were used to confirm results found in the initial analysis.

For all affective comments, the user’s heart rate, skin perfusion, and breath rate measures were analysed. For each of these, they were looked at in terms of average values, variance, and absolute change. It was believed that variance would be most sensitive to affect, as this is the value normally reported in psycho-physiological studies related to affect (Coles, 1983).

For each analysis, the results of the SPANOVA are reported, with results summarised in a table. Post-hoc analyses were performed whenever significant results were found. A summary of all primary aim results (initial and follow-up) is provided in section 8.3.5.c on page 195.

a. Initial Analysis

The results in this section are grouped according to the data reduction method – averages, variance, and absolute change. As a reminder, for the initial analysis only debrief comments (and associated physiological data) which did not have another affective comment made within the 25 seconds leading up to or following were analysed.

Averages Analysis

The SPANOVA performed on heart rate, skin perfusion, and breath rate averages resulted in no significant results, as summarised in Table 8.8.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Main Effects</th>
<th>Interaction Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>F(4, 57)=1.590, p&gt;0.05</td>
<td>F(4, 57)=1.590, p&gt;0.05</td>
</tr>
<tr>
<td>Skin Perfusion</td>
<td>F(4, 61)=0.678, p&gt;0.05</td>
<td>F(4, 61)=0.373, p&gt;0.05</td>
</tr>
<tr>
<td>Breath Rate</td>
<td>F(4, 49)=2.423, p&gt;0.05</td>
<td>F(4, 49)=2.157, p&gt;0.05</td>
</tr>
</tbody>
</table>

Despite no significant results, post-hoc tests were performed on Breath Rate as the main effects were found to be near significant (p=0.061) and it was considered prudent to explore this further, particularly because of the large F values being dealt with. A one-
way repeated measures ANOVA was used to compare the breath rate over time for both affective and non-affective conditions independently. There were no significant effects found for the non-affective condition with respect to breath rate \[ F(4,25)=1.427, \ p>0.05 \]. For the affective condition significant effects for breath rate were observed \[ F(4,26)=4.378, \ p<0.01, \ \eta^2=0.402 \]. This \( \eta^2 \) represents a large multivariate effect size according to Cohen’s criteria (Cohen, 1988). This means that the significant effect found is a true effect, and not an error. A paired t-test was then used to compare sequential intervals, revealing a significant increase between intervals 3 and 4 \[ t(30)=-3.534, \ p<0.01 \]. These intervals correspond to the time the comment made, and the ten seconds following the comment. No other significant effects were found.

**Figure 8.9: Heart Rate Averages**

No significant effects were noted.
Figure 8.10: Skin Perfusion Averages
No significant effects were noted.
* TPU = Tissue Perfusion Units, and is the unit of measure for skin perfusion

Variance Analysis
A SPANOVA performed on variance data revealed significant interaction effects for heart rate and skin perfusion (\([F (4, 57)=3.048, p<0.05]\) and \([F (4, 61)=4.424, p<0.01]\) respectively). A summary of all results are shown in Table 8.9.
Table 8.9: Variance SPANOVA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Main Effects</th>
<th>Interaction Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>F(4, 57)=2.263, p&gt;0.05</td>
<td>F(4, 57)=3.048, p&lt;0.05</td>
</tr>
<tr>
<td>Skin Perfusion</td>
<td>F(4, 61)=1.547, p&gt;0.05</td>
<td>F(4, 61)=4.424, p&lt;0.01</td>
</tr>
<tr>
<td>Breath Rate</td>
<td>F(4, 49)=0.891, p&gt;0.05</td>
<td>F(4, 49)=0.284, p&gt;0.05</td>
</tr>
</tbody>
</table>

Post-hoc tests were performed for heart rate and skin perfusion interaction effects. One-way repeated measures ANOVAs were performed to investigate effects of the interval (time) within either the affective or non-affective conditions and further paired t-tests were used if any significant effects were found through these ANOVAs. One-way between groups ANOVAs were used to investigate any differences between the same intervals for the two conditions.

Heart rate variance showed no effects with regards to time for the non-affective condition [F(4,27)=0.283, p>0.05]. However, the affective condition showed significant effects for the interval [F(4,27)=4.016, p<0.05, \eta^2=0.373]. Paired t-tests performed on the affective condition revealed that there was a significant increase in heart rate at the time of affective experience [t(30)=-4.150, p<0.01]. The one-way between groups ANOVA showed that heart rate in the affective condition was significantly lower than in the non-affective condition in the second time interval [F(1)=4.294, p<0.05] and significantly higher for the third interval [F(1)=6.560, p<0.05]. There were no other differences found between the conditions for heart rate.

The non-affective condition for skin perfusion showed no effects for interval [F(4,30)=0.477, p>0.05] whereas a significant effect was found for the affective condition [F(4,28)=4.954, p<0.01, \eta^2=0.414]. Performing paired t-tests on the affective condition revealed a significant increase in skin perfusion from the second to third intervals [t(31)=-2.850, p<0.01]. The one-way between groups ANOVA showed no significant differences between the conditions.
Figure 8.12: Heart Rate Variance

Significant effects (p<0.01) within a condition are indicated by a solid arrow. Significant effects (p<0.05) between the two conditions are indicated by a dashed arrow.

Figure 8.13: Skin Perfusion Variance

Significant effects (p<0.01) within a condition are indicated by a solid arrow.
Absolute Change Analysis

The mixed between-within subjects ANOVA revealed significant interaction effects for heart rate \([F (4, 61) = 2.950, p < 0.05]\) and for skin perfusion \([F (4, 64) = 3.572, p < 0.05]\). Breath rate did not show any effects. A summary of results, with significant effects highlighted, is shown in Table 8.10.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Main Effects</th>
<th>Interaction Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>(F (4, 61) = 2.290, p &gt; 0.05)</td>
<td>(F (4, 61) = 2.950, p &lt; 0.05)</td>
</tr>
<tr>
<td>Skin Perfusion</td>
<td>(F (4, 64) = 1.627, p &gt; 0.05)</td>
<td>(F (4, 64) = 3.572, p &lt; 0.05)</td>
</tr>
<tr>
<td>Breath Rate</td>
<td>(F (4, 55) = 0.513, p &gt; 0.05)</td>
<td>(F (4, 55) = 0.400, p &gt; 0.05)</td>
</tr>
</tbody>
</table>

Post-hoc one-way repeated measures ANOVAs were performed on heart rate and skin perfusion to investigate effects of the interval within both the affective or non-affective conditions. Paired t-tests were used if any significant effects were found through these ANOVAs. One-way between groups ANOVAs were used to investigate any differences at between respective intervals of the two conditions for these variables.

The one-way repeated measures ANOVA performed on heart rate showed no effects with respect to the interval for the non-affective condition \([F (4, 29) = 0.418, p > 0.05]\).
However, the affect condition showed significant effects between intervals \( F(4,29)=4.109, p<0.01, \eta^2=0.362 \). Paired t-tests performed on the affective condition revealed that there was an initial decrease in heart rate between intervals one and two \( t(32)=2.301, p<0.05 \), followed by a significant increase in heart rate at the time of affective response – between intervals two and three \( t(32)=-3.867, p<0.01 \).

The one-way between groups ANOVA showed that the affective condition was significantly lower than the neutral condition during the second interval \( F(1)=8.657, p<0.05 \). No other differences were noted between the affective and non-affective conditions.

Skin perfusion showed no effects for time in the non-affective condition \( F(4,33)=0.837, p>0.05 \), whereas a significant effect was found for the affective condition \( F(4,28)=3.946, p<0.05, \eta^2=0.360 \). Performing paired t-tests on the affective condition revealed a significant increase in skin perfusion between the second and third intervals (from just before the affective experience to the time of the affective experience) \( t(31)=-2.125, p<0.05 \). There was also a near significant decrease observed between intervals four and five of the affective condition \( t(31)=1.982, p>0.05 (p=0.056) \). The one-way between groups ANOVA showed that the affective condition was higher than the non-affective condition during the third interval, but this did not reach significance \( F(1)=3.427, p>0.05 (p=0.068) \). This is noted because it does appear that there would be a significant difference according to Figure 8.16,
Figure 8.15: Heart Rate Absolute Change

Significant effects (p<0.01) within a condition are indicated by a solid arrow. Significant effects (p<0.05) between the two conditions are indicated by a dashed arrow.

Figure 8.16: Skin Perfusion Absolute Change

Significant effects (p<0.01) within a condition are indicated by a solid arrow. Although it appears as though there would be a significant effect between the affective and non-affective conditions in interval 3, this was not supported by the statistical analysis.
b. Follow Up Analyses

The initial analysis reported in section 8.3.5.a only addressed comments that did not have another affect-related comment made in the 25 seconds leading up to or following. This was done to ensure that there were no additive effects or influences from adjacent affective experiences. However, this meant that the analysis was limited in power (n ~ 55). As such, it was considered useful to perform follow up analyses where trends leading up to an affective comment were analysed (regardless of whether or not another affective comment was made after the comment of interest), and trends following an affective comment were analysed (regardless of whether or not another affect comment was made immediately before the comment of interest). This increases the power by including comments that did not meet the stringent requirements of the initial analysis. Refer to Figure 8.7 on page 178 and related discussion for a more detailed explanation.

The results in this section are grouped according to the data reduction method – averages, variance, and absolute change. Results that differ from the initial analysis are emphasised. Results that are the same as the initial analysis are reported, but not elaborated on.
Averages Analysis

Results for the SPANOVA are summarised in Table 8.11. No significant effects were found.

Table 8.11: Averages SPANOVA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time Frame (with respect to comment made)</th>
<th>Main Effects</th>
<th>Interaction Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>Leading up to</td>
<td>F(2, 98)=0.181, p&gt;0.05</td>
<td>F(2,98)=0.532, p&gt;0.05</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>F(2,95)=0.567, p&gt;0.05</td>
<td>F(2,95)=0.784, p&gt;0.05</td>
</tr>
<tr>
<td>Skin Perfusion</td>
<td>Leading up to</td>
<td>F(2, 103)=0.442, p&gt;0.05</td>
<td>F(2,103)=0.117, p&gt;0.05</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>F(2,102)=0.809, p&gt;0.05</td>
<td>F(2,102)=0.980, p&gt;0.05</td>
</tr>
<tr>
<td>Breath Rate</td>
<td>Leading up to</td>
<td>F(2, 94)=1.494, p&gt;0.05</td>
<td>F(2,94)=2.076, p&gt;0.05</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>F(2,92)=3.039, p&gt;0.05</td>
<td>F(2,98)=1.325, p&gt;0.05</td>
</tr>
</tbody>
</table>

Variance Analysis

Results for the SPANOVA are summarised in Table 8.12.

Table 8.12: Variance Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Semi-Pure Type</th>
<th>Main Effects</th>
<th>Interaction Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>Leading up to</td>
<td>F(2, 98)=8.391, p&lt;0.01</td>
<td>F(2,98)=12.382, p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>F(2,95)=0.240, p&gt;0.05</td>
<td>F(2,95)=0.569, p&gt;0.05</td>
</tr>
<tr>
<td>Skin Perfusion</td>
<td>Leading up to</td>
<td>F(2, 103)=2.162, p&gt;0.05</td>
<td>F(2,103)=3.185, p&lt;0.05</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>F(2,102)=0.199, p&gt;0.05</td>
<td>F(2,102)=1.766, p&gt;0.05</td>
</tr>
<tr>
<td>Breath Rate</td>
<td>Leading up to</td>
<td>F(2, 94)=0.055, p&gt;0.05</td>
<td>F(2,94)=1.197, p&gt;0.05</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>F(2,92)=1.534, p&gt;0.05</td>
<td>F(2,98)=0.725, p&gt;0.05</td>
</tr>
</tbody>
</table>

In the post-hoc analysis of heart rate variance, it was noted that there was a significant decrease between intervals one and two \( t(57)=2.539, p<0.05 \), and a significant increase between intervals two and three \( t(57)=-6.110, p<0.01 \). The initial analysis only revealed the increase between intervals two and three were noted, with the difference between intervals one and two in the not reaching significance \( t(30)=1.878, p>0.05 \) (\( p=0.070 \)). However, the graph corresponding to the initial analysis for heart rate variance (Figure 8.12) does appear to show a decrease between intervals one and two for the affective condition. Therefore, this follow-up analysis seems to confirm that there is a decrease in heart rate variance between intervals one and two.
The follow-up analyses also confirmed the significant difference between affective and non-affective conditions for heart rate variance during the second interval (with the affective condition being significantly lower) \( F(1)=7.836, p<0.01 \). In addition, this analysis also revealed a significant difference in the third interval, where the affective condition showed a significantly higher heart rate variance than the non-affective condition \( F(1)=6.427, p<0.05 \). This was not found in the initial analysis.

The follow-up analysis found a significant increase in skin perfusion variance between intervals one and two for the affect condition \( t(56)=-2.263, p<0.05 \). By contrast, the initial analysis of skin perfusion variance showed a significant increase between intervals two and three. This difference is attributed to the follow-up analysis ‘negatively skewing’ the significant result to an earlier interval, as shown in Figure 8.18.

![Figure 8.18: Differences between Initial and Follow-up Analyses for Skin Perfusion](image)

**Absolute Change Analysis**

Results for the SPANOVA are summarised in Table 8.13.
### Table 8.13: Absolute Change Analysis

**Significant effects are highlighted**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Semi-Pure Type</th>
<th>Main Effects</th>
<th>Interaction Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>Leading up to</td>
<td>$F(2, 100)=6.229, p&lt;0.05$</td>
<td>$F(2, 100)=9.160, p&lt;0.01$</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>$F(2, 95)=0.856, p&gt;0.05$</td>
<td>$F(2, 95)=1.262, p&gt;0.05$</td>
</tr>
<tr>
<td>Skin Perfusion</td>
<td>Leading up to</td>
<td>$F(2, 98)=1.722, p&gt;0.05$</td>
<td>$F(2, 98)=3.821, p&lt;0.05$</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>$F(2,102)=0.925, p&gt;0.05$</td>
<td>$F(2,102)=1.501, p&gt;0.05$</td>
</tr>
<tr>
<td>Breath Rate</td>
<td>Leading up to</td>
<td>$F(2, 90)=0.044, p&gt;0.05$</td>
<td>$F(2, 90)=0.252, p&gt;0.05$</td>
</tr>
<tr>
<td></td>
<td>Following</td>
<td>$F(2, 89)=0.889, p&gt;0.05$</td>
<td>$F(2,89)=0.471, p&gt;0.05$</td>
</tr>
</tbody>
</table>

The follow-up analysis for heart rate absolute changes revealed the same results as the initial analysis.

However, an increase in absolute change in skin perfusion was observed between the first and second intervals for the affective condition [$t(53)=-2.280, p<0.05$], and there was no significant change observed between intervals two and three despite the initial analysis finding a difference between these intervals. This difference is attributed to negative skewing of skin perfusion data as discussed previously, and shown in Figure 8.18.

### c. Primary Results Summary

The primary aim of this study was to determine whether psycho-physiology or cued-recall debrief can be used as an evaluation method for user affect. The results found that there were significant physiological changes associated with affective comments made during the cued-recall debrief. No changes were found in physiological data when the user made a neutral comment. This confirms that both psycho-physiological methods and cued-recall debrief could be used as a method to evaluate user affect.

Specific trends that were observed included:

- There is a significant drop in heart rate variance and absolute change (i.e. it levels out) just before an affective comment. This is followed by a significant increase in variance and absolute change at the time of the affective comment.
- Skin perfusion variance and absolute change increased leading up to an affective comment. This was mirrored by a gradual decrease following an affective comment.
• There is a significant increase in the average breath rate immediately following an affective comment. This may be attributed to the user neglecting to breathe at the time of the affect.

The initial and follow-up analyses tended to show the same results, with some minor exceptions noted in the follow-up analysis section (8.3.5.b). These minor exceptions do not change interpretations of the effects noted above.

8.3.6. Results i Targeting Secondary Study Aims

The secondary aims of the study seek to determine the limitations of psycho-physiology as a method to evaluate affect, and are only applicable if psycho-physiology can be used for this purpose. The secondary study aims to determine what, if any, effects the following variables have on user physiological characteristics in response to affect:

• Affective valence: Is it possible for physiological measures to distinguish between positive and negative affect?
• Type of system used (i.e. game type): Do the physiological characteristics differ based on the type of game played?
• System preferences (i.e. game preference): Do the physiological characteristics differ based on user preference for a certain game?

a. Affective Valence

Studies performed by Healey and Picard (1998) have indicated that physiological measures cannot differentiate between positive and negative affective experiences. However, as valence data was recorded as part of this study (see 8.3.4) there was the opportunity to confirm or refute the findings by Healey and Picard. This is particularly important as Healey and Picard used an actress, and did not test this finding in the context of system use.

There were no significant effects found when comparing positive and negative affective experiences. This indicates that the psycho-physiological measures, as used within this study, were unable to differentiate between a user’s positive and negative affective reactions during system use. This confirms the results found by Healey and Picard (1998).
b. **Game Type**

The study used two different types of computer games – one action-oriented and one puzzle-style. These place different types of demands on the user, with one being more reaction based and the other being more cognitively based. The effects that this would have on the psycho-physiological recordings had not been explored.

There were no significant effects when comparing different types of games. This suggests that affect and related physiological change is independent of the system. That is, the same physiological changes will occur for a given user affective experience.

c. **Game Preference**

Brown (1988) and Malone and Lepper (1987) have each stated that if a person wants to or is motivated to use a certain system then they will become more engaged with the system. It was hypothesised that the degree of engagement may alter the physiological measurements related to affective experiences. Following the cued-recall debrief session, participants were asked to rate the computer game on a scale of -5 to +5. It was assumed that the higher the rating the more engaged the user was with that game.

When comparing preferred versus non-preferred games, skin perfusion variance did reveal an interaction effect between game preference and time for the affective condition \[F(4, 27)=2.721, p<0.05\], indicating that game preference does alter the physiological characteristics associated with an affective experience. However, post-hoc tests did not subsequently reveal any significant differences between game preferences. This could be due to the limited power afforded to the statistical analysis.

These mixed results mean that user preference may have an effect on the physiological effects associated with an affective experience. However, it should be noted that the game ratings may not be indicative of individual motivation for this study. Each participant did receive financial remuneration, and this external motivator may have created a situation where the participant was equally motivated to play any system (i.e. they will do anything provided they are paid). This may have masked any effects that could have been observed through physiological measures and motivation. Further study into the engagement of the use and physiological responses is necessary.
8.3.7. Summary of Results

The following points summarise all results obtained within this study. These are provided for ease of reference.

- Neutral comments were not related to any change in physiological variable.
- Heart rate variance and absolute change show a significant decrease just prior to the affective comment, followed by a significant increase at the time of the affective comment. There is a gradual (non-significant) decrease in heart rate variance and absolute change following an affective comment, as heart rate variance and absolute change return to normal levels.
- Skin perfusion variance and absolute changes increase significantly leading up to an affective comment. Though not statistically significant, there appears to be a gradual trend for skin perfusion variance and absolute change following an affective comment (returning to normal levels).
- It is not possible to differentiate between negative and positive affective experiences based upon the physiological measures recorded.
- Type of game has no impact on user physiology in relation to affect.
- Game preference may impact a user’s physiological response to affect, perhaps due to user motivation (i.e. users would be more motivated to play games that they prefer, and increased motivation has previously been linked to increased affect – see section 6.2.2).

8.3.8. Qualitative Review of Methods and Measures

The discussion below is divided into five parts, each focussing on an individual aspect of this study. These are:

- A general discussion about the use of physiological measures in relation to affect and system use (a);
- A focused discussion identifying the relative merits of each physiological variable measured:
  - Heart rate (b);
  - Skin perfusion (c);
  - Breath rate (d);
A focused discussion on the relative merits of the cued-recall debrief as a method to evaluate affect (e)

a. Physiological Measures (General)

Heart rate, skin perfusion, and breath rate were chosen as the physiological variables to monitor because previous studies have indicated that there is a link between these variables and affective experiences.

Each of these measures was considered non-obtrusive to the participant’s behaviour. This is necessary as any change in user behaviour may also influence user affect (Fishbein and Ajzen, 1975). Participants noted that after initial setup, they tended to forget that the physiological monitors were still on them. This is a testament to the fact that the equipment chosen was non-intrusive to their behaviour and did not alter their experience.

Three methods of data reduction were used in this study to analyse the physiological data (variance, absolute changes, and averages). Of these three data reduction methods, Coles (1983) had suggested that variance is the best option to use during psycho-physiological analysis. This suggestion was confirmed by this study as variance revealed the greatest number of differences in physiological variables as they relate to user affective experiences. Absolute changes were also sensitive to user affect; however, this is not as easy to calculate. Averages did not appear useful in the evaluation of affect. Only breath-rate was found to have any significant effect associated with averages, and for reasons discussed in section d, the method used within this study to measure breath rate is considered unreliable.

Therefore, it is recommended that variance analysis be used in future studies of this nature. Variance is sensitive to user affect, easier to calculate than absolute changes, and (in addition) is a commonly reported reduction method used in psycho-physiological literature (see Porges, Ackles, and Truax, 1983).

This study analysed physiological data in ten second intervals. This interval was set for several reasons:
• It was assumed that the participants will not always comment at the exact time of the affective (or non-affective) experience, and thus 10 seconds provides a small buffer zone within the data.

• It is not known whether the physiological changes related to affective experiences would be exhibited slightly before, at the same time, or slightly after the actual affective experience, as the theories relating affect and physiology do not offer this degree of precision (see section 7.4.2).

• Different physiological variables react at different rates to various stimuli, and the ten second buffer allows for some flexibility to identify a physiological change.

Some psycho-physiological studies have used smaller intervals than ten seconds – for example, some have studied heart rate on a beat-by-beat basis (see Andreassi, 1995). However, these studies sought to identify the physiological reactions of an individual when presented with a significant event, and were not trying to identify physiological responses where it is not possible to precisely identify the time when the ‘event’ occurred – where the event is an affective experience. As such, a beat-by-beat analysis does not afford the flexibility that a broader interval allows, and this flexibility is necessary when it is not possible to identify the exact time that an event happened.

b. Heart Rate via Pulse Oximeter

Heart rate was collected using a pulse oximeter for this study. The pulse oximeter was used because it was readily available to the researchers and software was available to input directly to a computer.

The requirements of this study and ensuing analysis did not require the accuracy or precision of the pulse oximeter, and similar results could have been observed using a typical heart rate monitor used in sport. Sport heart rate monitors are often watches that have a wireless connection to a chest strap and as such, they may even be preferable as it would not restrict movement in any way. Some sport heart rate monitors are also able to link to a computer to directly input data. The software associated with sport heart rate monitors tends to be much easier to use than the software for the pulse oximeter as noise filtering techniques are already integrated, whereas with the pulse oximeter this needed to be manually done.
Pulse oximeters also have the ability to measure pulse oxygenation levels. This functionality does not provide any benefit to the evaluation of system use, as the activity is not physically demanding enough to cause any variation in pulse oxygenation levels.

In short, the pulse oximeter was used in this study because it was readily available to the researchers and was able to input data directly to a computer. However, a sport heart rate monitor is possibly a better choice because it is easier to use, easier to setup, easier to record the data, and is less expensive.

c. **Skin Perfusion via Laser Doppler system**

The measure of skin perfusion was used in place of galvanic skin response. The equipment was readily available to the researchers and is considered more sensitive to physiological changes and less subject to participant movement than galvanic skin response (Shepherd and Öberg, 1990).

The skin perfusion measure provides a measure of the combined effects of blood pressure, sweat responses, and to a lesser degree muscle activation. Muscle activation primarily influences the reading through its impact upon localized blood pressure. This could cause issues in relation to activities performed during system use. For example, if an activity required constant or repetitive muscle activation the skin perfusion measure may be increased despite whole-body arterial blood pressure and sweat responses staying constant.

One major limiting factor to skin perfusion measurement is the associated costs. Laser-Doppler recording equipment is expensive, and it cannot be assumed that any industrial establishment would consider purchasing one for the purposes of affect evaluation (particularly as sport heart rate monitors are likely to work just as well). Galvanic skin response is not as cost prohibitive, and may be a better measure to use in place of skin perfusion. The sensitivity of galvanic skin response to user affect would need to be tested.

d. **Breath Rate via Nasal Thermistor**

Using a nasal thermistor to monitor breath rate was chosen because it was non-restrictive to the participant. Nasal thermistors have been used as a reliable method for
monitoring breath rate in sleep study participants (e.g. see Conduit et al., 2002; Russo-Magno et al.; 2001, or Hearne, 1982). These studies considered nasal thermistors to be an artefact free method of monitoring changes in respiratory rate. The actual measure of breath rate is obtained by using the temperature changes between exhaling and inhaling. The idea is that when a person exhales the air passing the thermistor is slightly warmer, and when they inhale the air passing the thermistor is slightly cooler.

During the study, the temperature change tracing caused by respiration was visible, and it was possible to see changes in breathing patterns. The conversion to breath rate was done using the Chart 4 software. The researcher was required to choose the noise filtering level for each participant, and whether each breath would be marked by consecutive minima or consecutive maxima. Though each of these should provide the same result, it was realized that some participants had more marked minima, and others had more marked maxima, and the appropriate selection was made for each participant. Figure 8.19 shows resultant graphs for both the nasal thermistor temperature tracing and the calculated breath rate.

![Nasal Thermistor Tracing and Conversion to Breath Rate](image)

**Figure 8.19:** Nasal thermistor raw output (temperature) and conversion to breath rate

At 15:10 the participant was breathing fast as shown by the short wave pattern in the tracing, and this was represented by an increase in the actual breath rate. At 15:35 the participant skipped a breath and this was shown by a sudden decrease in breath rate.

The main confounding variable in the use of nasal thermistors is that people’s breathing patterns are not uniform – that is they do not necessarily exhale and inhale the same degree each time. In Figure 8.19 there are times when the participant appears to hold
their breath (marked by ‘flat’ maximums), or take a deeply inhale, slightly exhale, then
inhale again (i.e. 15:35 in the above figure). Both of these cause the calculation of
breath rate to fluctuate greatly, and greatly limit the usefulness of nasal thermistor for
affect evaluation. This does not mean that breath rate cannot be used for affect
evaluation – simply that the nasal thermistor is not adequate. The nasal thermistor may
be suitable for sleep studies because they are normally not as concerned with very short
term changes, or sleep studies may not need to convert to breath rate, which is where the
problems occurred.

For breath rate to be considered as a method to evaluate affect it is necessary to have a
measure that is free from influences such as these in order to accurately identify
affective experiences (which may last only a short period of time). One alternative
measure is a full respirometer. This requires a participant to breath into a mouth piece
and is considered the most accurate of breath rate measures. This was considered too
restrictive on the participant, and may have impacted the other aspects of the study.
Another alternative is the use of a strain gauge to collect information about breath rate.
These gauges can be attached to an elasticized belt and secured either around the chest
or around the neck. When the participant breathes in, the gauge has an increased
negative pressure (i.e. belt is expanding) and this can be monitored effectively. These
may offer an alternative to nasal thermistors in measuring breath rate.

e. Cued-Recall Debrief

The cued-recall debrief was able to elicit information relating to affective experiences.
This information can be reliably matched to the current system state or progress,
meaning it is possible to identify what characteristics, behaviours, or elements of the
system created the user’s affective experience.

The cued-recall debrief method allows for greater insight into affective experiences than
psycho-physiology because it is possible to elicit specific information about the
experience – whether it was positive, negative, and can provide detail regarding the
exact nature of the affective experience. One shortcoming of this technique is that it
effectively doubles the evaluation session time as the participant must first perform all
activities, and then go through the debrief.
One issue experienced in relation to this is that comments elicited during the debrief relating to affective experiences tended to be vague in their time specification (e.g. one participant is quoted as saying “Somewhere around here I started to get excited”). This may cause problems matching the comment the system characteristics that created the affect. It is possible to use clues within the video to provide a better estimate of when the affective experience actually occurred. These clues may relate to the participant’s body language, comments that the participant made during the evaluation session (not the debrief), or it may be possible to use personal judgement and identify some system behaviour or characteristic that was likely related to the affect experienced.

8.3.9. Conclusions

The primary aim of this study was to determine if psycho-physiology or cued-recall debrief can be used as a method to evaluate user affect during product use. Since neither of these methods had been used for this purpose previously, there was no established benchmark that could be used. As such, this study relied on ‘self confirmation’ – that is, if a user made a debrief comment about an affective experience then this should be associated with some physiological change. This effectively means that the study required both psycho-physiology and cued-recall debrief to be capable of evaluating user affect. It has been established that there are physiological correlates to affective experiences in people (see Gale and Edwards, 1983), and cued-recall has been reported to be capable of eliciting comments about user affect (Omodei, 2002). What was unknown was whether the context of system use would influence the capability of either method.

This study confirmed that both cued-recall debrief and physiological measures are capable of being used to evaluate user affective experiences.

Cued-recall debrief is capable of identifying specific affects experienced. The limitation of cued-recall is that it relies on the user to report their affective experience, and the user may neglect to comment on or not realise they are having an affective experience. As such, cued-recall debrief will elicit information about specific affects, but there is the possibility that some affects will go unnoticed. Cued-recall could be used by itself effectively, relying on the subject to identify each affect experienced.
This information can then be given to system developers, who could then modify the system with the aim to further increase the positive and decrease the negative affective experiences.

The use of physiological measures can objectively determine when a user is having an affective experience. This is marked by changes in heart rate and skin perfusion variance. However, physiological measures cannot differentiate between different affects experienced, and cannot even discern whether the affect was positive or negative. Therefore, physiological measures can only indicate when a user is experiencing affect, and do not provide any insight regarding the nature of this affect. This limits the usefulness of physiological measures by itself – not being able to determine if the user is having a positive or negative experience limits the feedback able to be given to the developers.

Combined, physiological measures and cued-recall debrief appear to offer an optimal evaluation method. There is the objective indication of when a user is experiencing affect through physiological measures, which means that all affects are noticed. The cued-recall comments will identify the specific affects, which overcomes the limitation of physiological measures alone. Ideally, the evaluator of a system would know exactly when a user experienced affect from the physiological measures taken during use, and would prompt the user at these times during the debrief.

Despite appearing to be an optimal affect evaluation method, further research still needs to be performed. The method used in this study to process physiological signals took a great deal of time and effort. Using this method it would not be possible for the evaluator to enter into the debrief knowing when the user experienced affect (and therefore they would not know when to prompt the user). A more efficient method of processing physiological signals needs to be developed. Ideally, this would occur concurrently such that immediately after use the user would undergo the cued-recall debrief.

It is also necessary to study the generalisability of results. Within this study two very different computer games were played. The prime aim of games has been said to be the ability to create appropriate affective experience in users (Csikszentmihaly, 1992), and
therefore the use of computer games may actually heighten the affect experienced by a user. It is necessary to determine if physiological measures would be sensitive enough to user affect experienced during the use of less non-performance-based technologies (recall Figure 5.2, page 89). The two games, though very different, only allow results to be generalised to pure non-performance-based technologies, and may be limited only to computer games. Additional studies need to be performed with other types of systems to confirm this.

Another area of further work is to determine if other less medically-oriented physiological monitoring tools could provide the same results. One tool – sport heart rate monitors – are of primary interest. These are inexpensive, easily accessible, have software to download data to a computer, and are much easier to use than the medically based pulse-oximeter used in this study. Only one physiological measure is necessary provided it can indicate when a user is experiencing affect (i.e. there is no need to measure heart rate and skin perfusion). Further studies should be performed to test sport heart rate monitors for affect evaluation, and if suitable then these are recommended for use.

Finally, both psycho-physiology and cued-recall debrief need to be tested in an industrial setting. This study was conducted in a research environment, where there are resources often not available to industry. Although it appears that there are no exemplary resources or expertise required it is necessary to confirm this. It is also necessary to ensure that each method fits within the constraints imposed by industry – such as time-constraints for product development. Any evaluation method used must provide timely and useful input into the development process.

Both methods – physiological measurement or cued-recall debrief – appear to be viable for immediate industry uptake, subject to the limitations and necessary research specified. Sport heart rate monitors, if found valid for affect evaluation, are inexpensive and easy to use. Cued-recall debrief method does not require any specialised equipment. However, experience in conducting debriefs will improve debrief results; although useful information about user affect can still be obtained by an inexperienced debrief facilitator.
8.4. Chapter Summary

To validate a requirement there must be some way to measure and evaluate a variable that is directly associated with that requirement. Past QIU validation methods do not address affect, as discussed in Chapter 7.

Three approaches to evaluate affect in the context of system use were proposed (self-tracking, psycho-physiology, and cued-recall debrief), and this chapter described two studies that investigated their viability. Specifically, this chapter:

- Discussed the aims of affect evaluation. The ideal aim is to have an objective identification of the exact affective experience. However, it would be sufficient to have an objective indication of when an affective experience is occurring, and/or to have a subjective identification of the exact affective experience.
- Participant self-tracking was studied, and was not found to be a viable method to evaluate user affective experiences as it interrupted the behaviour of the participant and caused changes in their actual affective experiences.
- A study was performed to assess the potential of psycho-physiology measures and the cued-recall debrief as method to evaluating user affective experiences. Results found:
  - Heart rate and skin perfusion variance are sensitive to user affect, but cannot differentiate between specific affects. Breath rate was not found to be of use.
  - Cued-recall debrief can be used to identify specific affective experiences, but relies on the user to identify when they had these affective experiences.
  - Combined, physiology and cued-recall appear to offer an optimal evaluation method as there is the objective indication of when an affective experience is occurring (through physiology) and the identification of the exact affect experienced (through cued-recall).
The following chapter (Chapter 9) provides a review and overall discussion of this thesis, its aims, and how the research process addressed each of these aims. Future research projects that can extend this research within this thesis are also described.
Chapter 9  Review and Discussion of Thesis

The importance of user affect and emotion within the context of system use is increasing. Users are now demanding systems that are fun, enjoyable, or otherwise pleasurable whereas in the past these were considered ‘added bonuses’ – nice to have, but not a necessary requirement. Why this is happening is not entirely clear. It is possible that users now know what technology can do and are demanding the most from it. Alternatively, developers may be driving this trend as they realise there is little room for performance to improve, and therefore for their product to stand out they are designing to enhance user affective experiences. Regardless, understanding how to develop for and evaluate affect within the context of product use is becoming important.

This thesis began by exploring the theoretical aspects related to what affect actually is, how it is different from emotion and satisfaction, and how it relates to system quality – in particular quality in use. This thesis then provided some practical guidance on how to design to generate positive affect, and concluded by providing methods that can validate system requirements related to user affect.

This chapter provides an overall summary of this thesis in section 9.1, with emphasis on how the research objectives and questions were addressed. Possible future research directions related to affect and system quality are presented in 9.2. Section 9.3 evaluates the research process used within this thesis, and section 9.4 provides a final concluding comment to this thesis.

9.1. Review of Research Objectives

This thesis had the primary objective of identifying the importance of affect with respect to quality in use, and to suggest how affect can be addressed and validated within the context of system design. To achieve this, three specific research questions were posed in section 1.3:
1. Does affect impact system quality? If so, how?
2. What product design characteristics can aid in the creation of affective experiences in the user?
3. What is an effective method to validate affect as part of quality in use?

System quality was introduced in Chapter 2, with the emphasis on quality in use. This chapter highlighted the importance of, the factors that contribute to, and the relationship between the development process and quality in use. Figure 2.4 (page 18) provides an overview of what quality in use entails – namely effectiveness, productivity, safety, and satisfaction. However, sections 2.7 and 2.8 noted that these do not necessarily represent the user’s perception of quality (i.e. quality in use) for non-performance-based technologies. System quality deals with the “the ability of a system to achieve stated and implied needs of users”, and if those needs include affective experiences then these should be addressed within the framework of system quality, as this will influence quality in use experienced by the user.

Further discussion about what affect actually is, and how it relates to the design of a system revealed that affect was subject to a variety of interpretations in the literature. For example, it has been used synonymously with emotion in HCI literature. However, reviewing psychological research related to affect and emotion revealed that the two are related, but not the same. “Emotion” tends to be a longer term feeling, and includes the cognitive processing of a large number of influences. “Affect”, on the other hand, is a shorter term feeling and not subject to other influences (see 3.1.3).

Affect has also been considered equivalent to or a part of user satisfaction, and this was explored at length in Chapter 4. The necessity to explore this relationship stemmed from the fact that if affect were included within the notion of satisfaction, then affect would already be addressed within quality in use through satisfaction. However, a linguistic study of “satisfaction” showed that this was not the case. The study found that satisfaction related to feelings experienced when striving to achieve or after the achievement of some goal, and were not affective experiences. Other researchers performing similar studies in parallel came to the same conclusions (e.g. Lindgaard and Dudek, 2003). It should be noted that affect may contribute to satisfaction if the user’s goal was to experience affect, but affect is not satisfaction.
Within this discussion of satisfaction, the notion of user experience was raised, and this was discussed in Chapter 5. This chapter argued that user experience is equivalent to quality in use. This meant that it is possible to apply research and ideas related to user experience to quality in use. Within Chapter 5, a model of the overall user experience as it relates to different types of systems (e.g. performance-based or non-performance-based) was proposed (see section 5.2). This model suggested that the user experience depends on the type of system being used. The user experience of performance-based technologies will primarily be based on performance- and productivity-related factors, and user experience of non-performance-based technologies will be derived primarily from affective factors. Between these two extremes, the proportion of user experience derived from performance and productivity factors or affective factors will vary, depending on the product being used.

These chapters (Chapter 2 through Chapter 5) addressed the first research question. Affect can influence system quality and quality in use, if an affective experience is a desired goal of the user. Affect may also detract from the quality in use, if it was not a goal of the user and has a negative influence on the achievement of other goals. Affect is theorised to impact quality in use differently depending on whether it is performance-based or non-performance-based, with affect having a greater impact on the quality in use of non-performance based technologies (see Figure 5.2, page 89).

The second research question was primarily addressed in Chapter 6, though background material presented in section 3.3 also contributed to understanding the relationship between product characteristics and affect. Within Chapter 6 three theories that were said to relate affect and design were discussed – flow, internally motivating interfaces, and usability. A web-based survey asking participants to rate individual elements of these theories as they contribute to positive affective experiences identified several product characteristics that could enhance user affect. These included product learnability, increased feedback from the product, curiosity about what the product will do, cooperation with other users, competition against other users, and improved aesthetics (see section 6.3.6 for a more detailed description). Thus, by incorporating these characteristics into a system, it should enhance user affect. However, as with all system requirements there are necessary trade-offs, and some of these affect-enhancing characteristics are at direct odds with performance and productivity characteristics. For
example, it is advised to make performance-based technologies behave as a user expects, and adding elements of ‘curiosity’ will detract from performance as the user doesn’t know if their action will achieve the desired outcome. On the other hand, in the context of non-performance-based technologies it may be useful to include elements of curiosity, despite this having performance implications.

The final question – what is an effective method to validate user affect – was the subject of discussion and study in Chapter 7 and Chapter 8. Chapter 7 discussed why it was important to validate affect, and reviewed several common quality in use evaluation methods with respect to their capability of validating affect-related requirements. Each method reviewed was limited for this purpose. However, this review allowed for the development of a set of affect validation requirements (section 7.3). Three possible affect validation methods were proposed: self-tracking of affect; psycho-physiology; and cued-recall debrief.

Chapter 8 reported on two studies related to these proposed validation methods. The first study explored self-tracking of affect, where it was found that the activity of self-tracking could adversely impact the ability of the user to become engaged with the system, and therefore user affect will be altered. The second study explored psycho-physiology and cued-recall debrief as methods to validate affect. Both show promise for the evaluation of user affect. Psycho-physiology was capable of objectively indicating when a user experienced affect, but was not able to identify the specific experience. On the other hand, cued-recall debrief could identify the specific affective experience, but relied on the user to indicate when these experiences occurred. Combined, these two methods appear to provide an optimal evaluation method for user affect.

In summary, this thesis identified that affect is an important consideration for many modern products, most notably non-performance-based technologies. As such, affect should be considered within the framework of system quality, as system quality deals with “the ability of a system to achieve stated and implied needs of users” – where the needs can include affect. Design guidance has been provided which identifies characteristics of products that have been linked with an increase in positive affect.
Two evaluation methods – psycho-physiology and cued-recall debrief – have been studied and show promise as validators of affect-related requirements.

**9.2. Future Research Directions**

By having a deeper understanding of the issues related to a particular concept (i.e. affect as it relates to system quality), one becomes more aware of research limitations which lead to additional research directions, and also aware of new research questions.

This thesis discussed and explored only non-performance-based technologies. This focus ensured that results from studies were due to affect only, and not other confounding variables. However, this means that the interaction effects between affect and other variables (e.g. productivity, performance) are still unknown. It would be of interest to perform research to study the interaction effects between affect and other possible intervening variables.

Related, this thesis used computer games for all user-based studies. This ensured a high level of engagement and the elicitation of many affective experiences making the study of affect easier (see section 6.1 for greater detail). However, this does limit the generalisability of results. Products between the extremes of performance and non-performance technologies may show some interesting and unexpected trends in relation to the contribution of performance or affect to quality in use (recall see Figure 5.2, page 89). Studying user affect during the use of other products is one area which deserves additional research.

Another limitation of this research was not addressing the user’s emotion and mood, despite these contributing to the overall experience the user ‘walks away with’ (recall sections 3.1.2 and 3.1.5.a on emotion and mood respectively). The justification for the focus on affect is contained within 3.2, though can be summarised by stating that affect was chosen as the focus of this thesis because it is more sensitive to the impacts of specific interactive elements, whereas emotion tends is less sensitive to the specific elements (quote from 3.2, first paragraph). However, research should be conducted to look at the complete user experience, specifically accounting for the contributions of affect, emotion, and mood.
One question raised during the conduct of this research relates to the effect of context on the relationship between affect, performance, and satisfaction and type of product used. Figure 5.2 only shows the relative performance-orientation of products, and it is thought that if a product is mandated for use versus discretionary use this will influence the relationship. It is hypothesised that the discretionary use of performance-based technologies will increase the importance of the affective components, and the mandated use of non-performance based technologies will decrease the importance of affective components. Figure 9.1 is a revision of Figure 5.2 and shows the ‘context of use’ axis.

Another identified research area relates to the validation of design characteristics that can create positive affective experiences in a user (presented in section 6.3.6). These characteristics are supported by theories (section 6.2) and survey results (section 6.3.5). Even though there is a strong basis for accepting these characteristics as capable of enhancing positive affect, it is necessary to test these in situ. Ideally, a product where it is possible to alter the design characteristics (i.e. increase or decrease the aesthetic appeal, degree of learnability, etc.) between trials could be tested to determine what, if any, effect these characteristics have on product use.

The study into methods that can evaluate user affect (section 8.3) revealed several areas of future exploration. One result from this study suggested that user physiological variables differed based on user preference for a certain game. This may make it possible to compare between products and objectively identify which product the user
prefers. Another area of study is to determine the capability of cued-recall debrief and psycho-physiology to evaluate user affect in products other than computer games. Computer games were chosen to study within this thesis because they were believed to create a greater number of affective experiences in a user, which means there is a greater opportunity to study affect. It is not clear whether other product types will be able to create a strong enough affect to be measured using physiological measures. A final area of study that was noted is the need to determine whether or not psycho-physiology and / or cued recall can be used within the time constraints in an industrial environment. A research program for this final area has been proposed; with research commencing early 2006 (see Bentley and Saliba, 2005).

9.3. Research Process Evaluation

Overall, the research process used was effective. Each identified research question was addressed, and the relationship between quality in use and affect was explored in detail.

The main strength of this research process was the adoption of many different research methods from many different fields. These included methods adopted from linguistics, naturalistic decision making, system quality, usability, requirements engineering, psycho-physiology, and psychology. The adoption of these methods truly shows the breadth of fields that contribute to HCI and system quality research. In relation to this thesis, the adoption of these many methods of study meant that the relationship between affect and system quality was studied from many different viewpoints, with each enhancing the understanding of affect as it relates to system quality.

One issue that was encountered during this thesis process was the immense growth of research activity related to emotion and affect in the context of human-computer interaction. This improved the quality of research conducted within this thesis as it was possible to draw on and compare with other research studies, allowing greater trust to be placed in results which were consistent with other studies. However, it also meant that it was difficult to maintain abreast of all current literature in the area. It also meant that as new information, tools, and methods became available (through research within this thesis or other publications), new ideas and new paths emerged. To account for this, the original thesis plan was continually reviewed and revised to ensure that it accounted for
these new ideas and paths. As such the formation of this thesis was, in part, reactive. It would have been ideal if the original thesis plan was followed from the beginning, but with any long-term research project one should expect and be prepared to react to new information as it becomes available. However, this thesis did have specific stages and key goals identified, and these did not change throughout the thesis process.

9.4. Final Comment

Affect has recently become a topic of discussion in many HCI related forums. With workshops being performed on ‘Fun’, panels discussing the relative merits of affect in various contexts, and an increasing body of literature on affect in HCI it is apparent that academics, researchers, and practitioners have accepted affect as important. However, “affect” is still loosely used, and may represent anything related to emotion or satisfaction. This can lead to work using similar terms but meaning different things, or work using different terminology but meaning the same thing. This will, in the least, cause problems with communicating knowledge to other researchers or practitioners. However, this may also lead to confusion and perhaps disrepute for the study of affect within the context of system quality. This thesis provides usable operational definitions for much of the affect-related terminology, and provides an outline of the different affect research areas (section 3.3).

While this is useful in itself, for affect to be truly understood within HCI it is necessary to address how it can be influenced by the design of a system, and how we can reliably and validly evaluate human affect. Without addressing these latter issues any discussion of affect within HCI is redundant. Work described in this thesis shows that there are specific characteristics of systems that do tend to be related to positive affective experiences. Further, human physiology and situated recall methods (e.g. cued-recall) can provide insight into these affective experiences and provide an avenue for both application and evaluation.

This thesis has presented a case for including affect as a contributor to system quality. Quality in use, as discussed in this thesis, is flexible as it is recognised that not all qualities will be important to all users or to all systems. This is true with affect as well
– there are some systems where affect will be a prime determinant of quality, and others where affect has little or no influence on quality.

At the beginning of Chapter 2 a short quote was given. Below is the complete version:

Quality -- you know what it is, yet you don't know what it is. But that's self-contradictory. But some things are better than others, that is, they have more quality. But when you try to say what the quality is, apart from the things that have it, it all goes poof! There's nothing to talk about. But if you can't say what quality is, how do you know what it is, or how do you know that it even exists? If no one knows what it is, then for all practical purposes it doesn't exist at all. But for all practical purposes it really does exist. What else are grades based on? Why else would people pay fortunes for some things and throw others in the trash pile? Obviously some things are better than others – but what's the "betterness"? -- So round and round you go, spinning mental wheels and nowhere finding anyplace to get traction.

What the hell is Quality? What is it?

Pirsig, 1984
(concluding paragraph, Chapter 15)

It is hoped that this thesis has shed some light in regards to what creates quality for users of some products …

Todd
References


ISO. (1999). ISO 13407: Human-centred design processes for interactive systems. ISO.


Appendix A: Quality and Quality Standards Summary

Bevan (2001) identifies specific sub-groups related to system quality. These groupings are shown in Table A.1. The guidelines (Table A.2) offer suggestions and ideas that may improve software quality. It is necessary for the developer to interpret and apply the guidelines as they see fit. The guidelines presented in the standards deal with a vast array of issues related to use, development process, hardware and software interface design, documentation, and related issues such as accessibility. Specifications (Table A.3) provide concrete ideas of how to design for a quality product. The ideas offered detail exactly how a system should behave, look, feel, etc. in a given context. Specifications offer the developer a pre-conceived design that only requires implementation and integration with the system being developed. Guidelines differ from this by also requiring the developer to interpret it and use it as a basis for a design.

<table>
<thead>
<tr>
<th>Area</th>
<th>Sub-Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality in Use</td>
<td>Product use in context</td>
<td>Addresses the quality of a product while being used in a particular context</td>
</tr>
<tr>
<td>Product Quality</td>
<td>Software Interface and Interaction</td>
<td>Supports the development of a software user interface by providing details (specifications), guidance on development, and evaluation criteria</td>
</tr>
<tr>
<td></td>
<td>Hardware Interface and Interaction</td>
<td>Supports the development of a hardware interface by providing details (specifications), guidance on development, and evaluation criteria</td>
</tr>
<tr>
<td></td>
<td>Documentation</td>
<td>Provides details of the process to development of support (user) documentation, and describes the user needs with respect to documentation</td>
</tr>
<tr>
<td>Process Quality</td>
<td>User-centred processes</td>
<td>Explains the activities, methods, and frameworks that can be used during development to support the creation of a quality product</td>
</tr>
<tr>
<td>Organisational Capability</td>
<td>Usability Capability</td>
<td>Provides the basis of system quality. This area of quality is the assessment of the actual ability of the organisation to perform ‘best-practice’ design activities. (i.e. are the resources available)</td>
</tr>
<tr>
<td>Standard</td>
<td>Title</td>
<td>Use In Context</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>ISO/IEC 9126</td>
<td>Software Engineering – Product Quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part 1: Quality Model</td>
<td>[✓]</td>
</tr>
<tr>
<td></td>
<td>Part 2: External Metrics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part 3: Internal Metrics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part 4: Quality in use metrics</td>
<td>[✓]</td>
</tr>
<tr>
<td>ISO 9241</td>
<td>Ergonomic requirements for office work with visual display terminals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part 1: General introduction</td>
<td>[✓]</td>
</tr>
<tr>
<td></td>
<td>Part 2: Guidance on task requirements</td>
<td>[✓]</td>
</tr>
<tr>
<td></td>
<td>Part 11: Guidance on usability</td>
<td>[✓]</td>
</tr>
<tr>
<td></td>
<td>Part 10, 12-17 (inclusive): Interface dialogue design</td>
<td>[✓]</td>
</tr>
<tr>
<td>ISO 14915</td>
<td>Software ergonomics for multimedia user interfaces</td>
<td></td>
</tr>
<tr>
<td>IEC TR 61997</td>
<td>Guidelines for the user interfaces in multimedia equipment for general purpose use</td>
<td></td>
</tr>
<tr>
<td>ISO 11064</td>
<td>Ergonomics design of control centres</td>
<td></td>
</tr>
<tr>
<td>ISO/IEC 18019</td>
<td>Guidelines for the design and preparation of software user documentation</td>
<td></td>
</tr>
<tr>
<td>ISO 13407</td>
<td>Human-centred design processes for interactive systems</td>
<td></td>
</tr>
<tr>
<td>ISO TR 16982</td>
<td>Usability methods supporting human centred design</td>
<td></td>
</tr>
<tr>
<td>ISO TR 18529</td>
<td>Ergonomics of human-system interaction – Human-centred lifecycle process descriptions</td>
<td></td>
</tr>
<tr>
<td>ISO 10075</td>
<td>Part 1: Ergonomic principles related to mental workload – general terms and definitions</td>
<td></td>
</tr>
<tr>
<td>ISO DTS 16071</td>
<td>Guidance on accessibility for human-computer interfaces</td>
<td></td>
</tr>
</tbody>
</table>
Table A.3: Summary of Software Quality Standards

These standards and their descriptions were provided by Bevan, 2001

<table>
<thead>
<tr>
<th>Standard</th>
<th>Title</th>
<th>Use In Context</th>
<th>Software Interface and Hardware Interface</th>
<th>Documentation</th>
<th>Development Process</th>
<th>Usability Capability</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 20282</td>
<td>Usability of Everyday Products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO/IEC 10741</td>
<td>Part 1: Dialogue interaction – Cursor control for text editing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO/IEC 11581</td>
<td>Icon symbols and functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO/IEC 18021</td>
<td>Information Technology – User interface for mobile tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO 9241</td>
<td>Parts 3-9 (inclusive) Ergonomic requirements for office work with visual display terminals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO 13406</td>
<td>Ergonomic requirements for work with visual displays based on flat panels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO/IEC 14754</td>
<td>Pen-based interfaces – Common Gestures for text editing with pen-based systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO 18789</td>
<td>Ergonomic requirements and measurement techniques for electronic visual displays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO/IEC 15910</td>
<td>Software user documentation process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ISO/IEC 14598</td>
<td>Information Technology – Evaluation of software products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Appendix B: Descriptions of the Human-Centred Design Process

Each stage of the human-centred design process, as provided by ISO 13407, is summarised in the following table. Detailed descriptions of the stage and activities that should be followed in each stage are available in ISO 13407 and ISO TR 18529. Combined, these are considered to be the minimal acceptable standard of human-centred development processes (Earthly, Sherwood-Jones, and Bevan, 2001).

<table>
<thead>
<tr>
<th>Stage</th>
<th>Summary Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan the Human-Centred Design Process</td>
<td>This stage is critical for the successful integration of humans into the development process. The plan will drive the remaining stages of development and will encourage an iterative design approach with constant feedback confirming and providing new input into the development strategy. Planning should involve consultation with stakeholders to gain an understanding of their objectives and the constraints that must be placed on the system. This includes identifying the target users and their capabilities. It should also include a strategy to involve the users throughout the development. If the plan is poorly conceived then often the user may be neglected during the development process. If this occurs, the development team may need to make assumptions regarding the context of use, user characteristics, and specific user requirements. It is very likely that some or all of these assumptions will be incorrect or incomplete; thereby the developed system will be based on an incomplete understanding of the overall requirements. These false assumptions may lay dormant, and may not manifest themselves until the system is being used in context. Time and resources would be necessary to correct the mistakes due to the false assumptions, ultimately leading to an increased cost of the system – assuming they could be fixed within imposed time and resource constraints at all.</td>
</tr>
<tr>
<td>Stage</td>
<td>Summary Description</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Identify the Context of Use</td>
<td>The context of use plays a primary role in determining what a system must be able to do and how it will do it. The context of use includes the attributes of the user that may influence the design of the system, the tasks that the system must be able to perform, the organisational needs, and technical and physical environmental considerations for the system. As such, the identification of the proper context of use often will place constraints on the final design of the system. Proper understanding of the context of use is critical for subsequent development stages. Extensive stakeholder input is important, particularly where the stakeholder has an intimate knowledge of the context being designed for. The context of use is relatively invariable for the design – particularly with respect to the organisation and environmental aspects. Throughout the lifecycle process the design team gains a greater understanding of the context, and with this increased understanding modifications to existing or the addition of new requirements may be necessary. By making sure that users are involved throughout the development cycle, the development team minimise the likelihood of significant changes to their understanding of the context of use.</td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td><strong>Summary Description</strong></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Specify user and organisational requirements</td>
<td>The stakeholders, who include organisational representatives and the targeted users, should assist in identifying the system requirements by identifying specific needs, wants, goals, and constraints with respect to the system including the identification of any risks associated with system use. The stakeholders should also define the role the system will have within the targeted context, perhaps identifying the system as integral to the overall context or as a minor support system. Requirements should be based upon all information including inherent constraints for the context, or those that are imposed by external sources (e.g. legislative constraints). Requirements may relate to the functionality required (i.e. what the system must do), or they may relate to quality aspects of the system (i.e. an efficient system, an easy to use system, etc.). With each requirement a viable measure to test whether the requirement is met should be set. Setting measures early ensures that there is a consistent understanding of the requirement between the development team, the users, the organisation, and other stakeholders. Stakeholders, in addition to providing information that forms the basis for the requirements, should assist in the prioritisation of the requirements, approve the measures, and ensure that the design team is conforming to statutory or legislative requirements. Understanding user and organisational requirements is one of the critical factors to the success of interactive systems (Maguire and Bevan, 2002). Benefits from a clear understanding of these requirements will include reductions in support and training costs, improved user satisfaction, and enhanced quality of work. Even so, successive iterations through the lifecycle process will see these requirements further refined as feedback is received through system evaluation, which will in turn refine the design.</td>
</tr>
<tr>
<td>Stage</td>
<td>Summary Description</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Produce design solutions</td>
<td>Producing design solutions relies heavily upon the skills and experience of the design team and their access to relevant information such as detailed requirements. The design team needs to collect and interpret all the information that they have available and make suitable choices and decisions regarding how to develop the system. The team will need to understand the task being performed, and appropriately allocate functions between the human and the technical system to take advantage. The use of low-fidelity prototypes allow the design team to communicate their ideas to the stakeholders for feedback, and also permit early evaluation which may identify shortcomings of the design before excessive resources are consumed on a sub-optimal design. Later during the lifecycle process, training and support for the user should be developed as appropriate to allow for smoother integration of the system into the organisational processes. Bass et al. (2003) indicate that design teams will often focus their resources on this stage, neglecting the other stages and aspects of the human-centred design process. Focussing on this stage means that the iterative nature of the design lifecycle has been neglected. Without the iterative process stakeholder involvement, refinement of ideas, refinement of requirements, and ultimately the refinement of design will often not occur.</td>
</tr>
<tr>
<td>Evaluate design against requirements</td>
<td>The evaluation of a system compares the designed attributes of a system to the system requirements and related measures. Ideally, the design will meet all requirements – meaning that the system is exactly as the users and stakeholders mandated. During the lifecycle process the evaluation stage identifies which requirements are and which are not being adequately met. The results of this evaluation will then feed back into the re-design of the system. Evaluation may also necessitate the refinement of sub-optimal requirements or may reveal new, previously unconsidered, requirements. The evaluation should take place in a context that simulates or is the actual context in which the system will be used. This will ensure that results from the evaluation are representative of the system in context, and is not an artefact of the environment used for the evaluation. This becomes increasingly important through subsequent iterations of the lifecycle. The evaluation process should not end when the system is implemented as it is ideal to continue evaluation of the system while it is being used in its’ true context.</td>
</tr>
</tbody>
</table>
Appendix C: Testing of Theories Relating Affect and Design

C.1 Web Based Survey

The online survey investigating the characteristics of computer games that contribute to user positive experience(s) while interacting with the system. The intent of this survey was to identify design characteristics that are correlated with positive affective experiences.

C.1.1 Introduction Page (Brief)

Investigating the Reasons why People Play Computer Games

Summary Information

Centre: SCHL, Swinburne University of Technology
Investigator: Todd Bentley (tbentley@it.swin.edu.au)
Supervised by: Lorraine Johnston (johnston@it.swin.edu.au)

The purpose of this study is to provide an initial look at what aspects of computer games people enjoy. To participate in this study, you are requested complete the following questionnaire that asks you to:

- Provide generic background information (age, gender);
- Rate up to 3 computer games that you enjoy on several scales;
- Provide any additional comments if you feel your reasons for enjoying a computer game have not been fully explained.

This questionnaire will take about 10 minutes to complete and is completely anonymous. Only data entered by yourself will be recorded. This study has been reviewed by, and received clearance through, the Human Research Ethics Committee at Swinburne University of Technology.

For more detail about this study (including the person to contact regarding the ethics of this study) please see the complete information page.

Proceed
Investigating the Reasons why People Play Computer Games

Information Form

Centre: SCHIL, Swinburne University of Technology

Investigator: Todd Bentley (tbentley@it.swin.edu.au)
Supervised by: Lorraine Johnston (ljohnston@it.swin.edu.au)

The purpose of this study is to provide an initial look at what aspects of computer games people enjoy. To participate in this study, you are requested complete the attached questionnaire that asks:

- Provide generic background information (age, gender) as this is important for interpretation of results;
- Rate up to 3 computer games that you enjoy on several scales;
- Provide any additional comments if you feel your reasons for enjoying a computer game have not been fully explained.

This questionnaire will take about 10 minutes to complete.

This study will allow the investigators to begin to understand what makes computer games enjoyable and/or engaging. Computer games have been targeted because it is unknown why people continue using many leisure technologies. And insight into a single technology is hoped to begin to address this issue. The results of this study will enable the investigators to provide a hypothesis of why people play computer games.

The questionnaire is completely anonymous. Results from the study will be published in an aggregate form in a suitable academic journal. You are free to withdraw from this study at any time. Should you have any questions regarding this study, feel free to e-mail the researchers at gamesresearch@it.swin.edu.au.

CONSENT: By completing this questionnaire you have agreed to participate in this research project, realising that complete anonymity will be maintained. You are not obliged to complete any part of this questionnaire, but please note that any data submitted cannot be removed.

This study has been reviewed by, and received clearance through, the Human Research Ethics Committee at Swinburne University of Technology. If you have any concerns regarding the ethical nature resulting from your participation in this study, please contact the Chair of the Human Research Ethics Committee:

The Chair
Human Research Ethics Committee
Swinburne University of Technology
P O Box 218
HAWTHORN, VIC. 3122
Phone: (03) 9214 6223

Proceed >>
Section A - Background Information

(Page 1 of 3)

This survey is intended to gain insight into the factors that encourage people to use certain leisure technologies, specifically computer games.

This questionnaire should take about 10 minutes of your time. All information you provide is totally anonymous, and no information recorded will be able to identify you. Your help would be greatly appreciated.

1. Age Range.
   - □ < 18
   - □ 18-25
   - □ 26-30
   - □ 31-35
   - □ 36-45
   - □ 46-55
   - □ 56-65
   - □ 66 +

2. Gender
   - □ Male
   - □ Female

3. Please describe your occupation in the space below.

4. How many different computer games would you consider you play regularly?
   - □ 0
   - □ 1-3
   - □ 4-8
   - □ 7-10
   - □ 10 +

5. How would you describe or classify the genre(s) (e.g. action, solitaire, simulator) of computer games that you prefer most?
   - □ Not Applicable - I have no preference

6. Where did you hear about this survey?

Continue to Section B  Reset Form
### Section B - Games

For this section, list a computer games that you play (or have played), and rank each of the statements with respect to how much you agree with that statement for this specific game.

**Game Name:**

How enjoyable do you find this computer game?

Approximately how much time do you spend playing this computer game?

<table>
<thead>
<tr>
<th></th>
<th>Highly Agree</th>
<th>Agree</th>
<th>Indifferent</th>
<th>Disagree</th>
<th>Highly Disagree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I enjoy this game because the information presented is not confusing.

I enjoy this computer game because I can try to beat my personal best.

I enjoy this computer game because it is clear what must be achieved to complete a level or win the game.

I enjoy this computer game because I do not need to devote much time and energy to learn how to play.

I enjoy this computer game because the computer responds appropriately to actions taken by others, the game, or myself.
<table>
<thead>
<tr>
<th></th>
<th>Highly Agree</th>
<th>Agree</th>
<th>Indifferent</th>
<th>Disagree</th>
<th>Highly Disagree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>I enjoy this computer game because it makes me forget about my worries and frustrations while I am playing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I enjoy this computer game because it lets me do what I want when I want.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I enjoy this computer game because I can quickly figure out how to achieve the objectives of the game or level.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I enjoy getting recognition from other people when I play the computer game well.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I enjoy this computer game because after I stop playing it I feel better about myself.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I enjoy this computer game because I can compete with others who have played it by trying to beat them or their scores.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>I enjoy this computer game because it makes time seem to go faster - hours can pass by in minutes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>I enjoy this computer game because I can consistently complete the game or levels with few errors.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>I enjoy this computer game because I am always wondering what will happen next, or what will occur if I take a specific action.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>I enjoy this computer game because I can identify with the situations and/or people presented in this computer game.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Highly Agree</td>
<td>Agree</td>
<td>Indifferent</td>
<td>Disagree</td>
<td>Highly Disagree</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------</td>
<td>-------</td>
<td>-------------</td>
<td>----------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>16</td>
<td>I enjoy being able to do things in this computer game that I can not do in everyday life.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>I enjoy this computer game because it is challenging, but within my ability to complete.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>I enjoy playing this game because I can play it with others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>I enjoy the recognition I get from the computer when I play the computer game well.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please feel free to make any additional comments regarding why you enjoy this game.

Submit This Form  Reset Form

**NOTE**: If you select "Submit This Form", you will be given the opportunity to describe the reasons you enjoy another game (up to 3 total) or finish the questionnaire. This will be game number.
C.1.5 Survey Appreciation Page

Thank you for completing this questionnaire.

The results of this survey will be posted at
http://www.it.swin.edu.au/schil/games/results/
when available.

If you would like to participate in future studies, or for more information about this study, please e-mail gamesresearch@it.swin.edu.au.

C.2 Initial Data Analysis

C.2.1 PCA Analysis (Initial)

<table>
<thead>
<tr>
<th></th>
<th>Kaiser-Meyer-Olkin Measure of Sampling Adequacy</th>
<th>$\text{KMO}$</th>
<th>Bartlett's Test of Sphericity</th>
<th>$\text{Chi-Square}$</th>
<th>$\text{df}$</th>
<th>$\text{Sig.}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.801</td>
<td></td>
<td>1725.894</td>
<td>171</td>
<td>.000</td>
</tr>
</tbody>
</table>
Initial component matrix extracted (below) reveals four individual components.

![Scree Plot](image)

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Recognition</td>
<td>0.703</td>
<td></td>
<td>-0.498</td>
<td></td>
</tr>
<tr>
<td>Recognition Other</td>
<td>0.642</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Competition</td>
<td>0.602</td>
<td>-0.306</td>
<td>-0.574</td>
<td></td>
</tr>
<tr>
<td>Self Awareness</td>
<td>0.586</td>
<td></td>
<td></td>
<td>-0.411</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>0.531</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Competition</td>
<td>0.523</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.509</td>
<td>-0.396</td>
<td>0.412</td>
<td></td>
</tr>
<tr>
<td>Distraction</td>
<td>0.500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.499</td>
<td>0.403</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.466</td>
<td></td>
<td></td>
<td>-0.458</td>
</tr>
<tr>
<td>Attribution</td>
<td>0.328</td>
<td>0.601</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curiosity</td>
<td></td>
<td>0.569</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>0.402</td>
<td>0.529</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goals</td>
<td>0.441</td>
<td>-0.441</td>
<td>0.441</td>
<td></td>
</tr>
<tr>
<td>Cooperation</td>
<td>0.468</td>
<td>-0.324</td>
<td>-0.587</td>
<td></td>
</tr>
<tr>
<td>Learnability</td>
<td>0.353</td>
<td>-0.454</td>
<td>0.475</td>
<td></td>
</tr>
<tr>
<td>Concentrate</td>
<td>0.445</td>
<td>-0.359</td>
<td>0.446</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>0.398</td>
<td></td>
<td></td>
<td>0.500</td>
</tr>
<tr>
<td>Feedback</td>
<td>0.376</td>
<td>0.310</td>
<td></td>
<td>0.392</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

a. 4 components extracted.
C.2.2  PCA Analysis (Final)

The final component matrix (when forced to use four factors) is shown below.

<table>
<thead>
<tr>
<th>Rotated Component Matrix$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
</tr>
<tr>
<td>Self Competition</td>
</tr>
<tr>
<td>Cooperation</td>
</tr>
<tr>
<td>Computer Recognition</td>
</tr>
<tr>
<td>Computer Competition</td>
</tr>
<tr>
<td>Recognition Other</td>
</tr>
<tr>
<td>Goals</td>
</tr>
<tr>
<td>Efficiency</td>
</tr>
<tr>
<td>Concentrate</td>
</tr>
<tr>
<td>Learnability</td>
</tr>
<tr>
<td>Effectiveness</td>
</tr>
<tr>
<td>Self Awareness</td>
</tr>
<tr>
<td>Time</td>
</tr>
<tr>
<td>Distraction</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Challenge</td>
</tr>
<tr>
<td>Feedback</td>
</tr>
<tr>
<td>Curiosity</td>
</tr>
<tr>
<td>Fantasy</td>
</tr>
<tr>
<td>Attribution</td>
</tr>
</tbody>
</table>

*Extraction Method: Principal Component Analysis.*

*Rotation Method: Varimax with Kaiser Normalization.*

*a. Rotation converged in 6 iterations.*
Appendix D: Participant Self Tracking

D.1 Option 1 Sample Scale

Scale Description: Placing avatars (emoticons) at regular thirty second intervals throughout the timeframe.

Figure D.1: Stereotypical Sample of Option 1

These are simulated results that were typical for the participants who completed this option. This scale has been shrunk to a single page width. The scale width was originally 70 cm.
D.2 Option 2 Sample Scale

Scale Description: Placing avatars (emoticons) at whenever a change affective experience occurs.

Please indicate any affective change that you experience by placing the most appropriate icon in the associated time slot. To begin, place the icon that best represents your current affective experience in the 0 minute slot.

Figure D.2: Stereotypical Sample of Option 2
These are simulated results that were typical for the participants who completed this option. This scale has been shrunk to a single page width. The scale width was originally 70 cm.
D.3 Option 3 Sample Scale

Scale Description: Continually reposition the slider to represent the current affective valence.

**Please continually** rate your current affective experience as either positive or negative on the following scale by moving the slider in the appropriate direction. Please try to remember to move the slider whenever your affect changes. Begin by placing the slider at the point representing your current affect.

![Scale Diagram]

**Figure D.3: Stereotypical Sample of Option 3**
The slider position would be recorded and later placed on a timeline.
D.4 Option 4 Sample Scale

Scale Description: Placing a verbal label representing current affect at regular thirty second intervals throughout the timeframe.

Please indicate the affective experience that you have had during the past 30 seconds by placing the most appropriate icon in the associated time slot. Please repeat this every 30 seconds. To begin, place the icon that best represents your current affective experience in the 0 minute slot.

<table>
<thead>
<tr>
<th>Neutral</th>
<th>Neutral</th>
<th>Excited</th>
<th>Surprised</th>
<th>Happy</th>
<th>Happy</th>
<th>Contented</th>
<th>Neutral</th>
<th>Neutral</th>
<th>Embarrassed</th>
<th>Surprised</th>
<th>Excited</th>
<th>Frustrated</th>
<th>Frustrated</th>
<th>Neutral</th>
<th>Frustrated</th>
<th>Happy</th>
<th>Amused</th>
<th>Amused</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 min</td>
<td>0.5 min</td>
<td>1 min</td>
<td>1.5 min</td>
<td>2 min</td>
<td>2.5 min</td>
<td>3 min</td>
<td>3.5 min</td>
<td>4 min</td>
<td>4.5 min</td>
<td>5 min</td>
<td>5.5 min</td>
<td>6 min</td>
<td>6.5 min</td>
<td>7 min</td>
<td>7.5 min</td>
<td>8 min</td>
<td>8.5 min</td>
<td>9 min</td>
<td>10 min</td>
</tr>
</tbody>
</table>

Figure D.4: Stereotypical Sample of Option 4

These are simulated results that were typical for the participants who completed this option. This scale has been shrunk to a single page width. The scale width was originally 70 cm.
D.5 Avatar Short List

The original avatar list was reduced to the following set based upon the results of the participant self-tracking study. This list was used during the cued-recall debrief portion of a subsequent study (see section 8.3).

The avatars below were removed from the original emoticon list.
Appendix E: Psycho-Physiology and Cued-recall Study

E.1 Participant Information Forms

**Physiology measures as an Indicator of Affect**

Information Form

HREC File Number:

Centre: SCHIL, Swinburne University of Technology
Investigator: Todd Bentley (tbentley@it.swin.edu.au)
Supervisory Investigator: Lorraine Johnston (ljohnston@it.swin.edu.au)

The aim of this project is to determine if physiological measures are able to indicate whether a person is experiencing an 'affective response' while using a software system. An affective response can include being engaged, feeling anxiety, surprised, having fun, and other similar elements. Currently affective responses are measured through subjective means – the user of a system will indicate after using the system if they have had fun, if they felt nervous, etc. This research project aims to provide an objective indicator that is able to be measured during system use, not following.

As a participant in this study, you would be asked to play two of the following computer games:

- A driving simulator
- A puzzle style game
- A strategy style game

You would be required to play each computer game for a period of 20 minutes. During this time you would have your heart rate, respiration, and skin perfusion (how 'flush' your skin is) monitored by physiological monitoring equipment, and you would be video taped. Following the 20 minute session, you would then be asked to watch the tape of yourself playing the game, and to provide feedback at times about what you were thinking and feeling while playing the game. You would then repeat this for a second computer game type.

Overall, this would require approximately 2 hours of your time. You would be provided with a $40 reimbursement to thank you should you decide to participate in this study. There are no anticipated risks during this study. All physiological monitoring equipment has been specifically designed for monitoring biological rhythms, and there is no undue risk to you as a participant.

All data collected through participation in this study will remain confidential. All participants will be identified throughout the study through the use of code numbers, and only the investigators listed above will know of your identity. Video taken of participants will not be shared in any way without prior consent of the participant involved.

You are free to withdraw from this study at any time. Results from this study are intended to be published in an aggregate form in a suitable academic journal and/or conference venue. If you have any questions regarding the project entitled Physiological Measures as an Indicator of Affect, you can direct them to Todd Bentley.
of the Swinburne Computer Human Interaction Laboratory (SCHIL) on telephone number 03 9214 8860 or alternatively via e-mail at tbentley@it.swin.edu.au.
E.2 Participant Consent Form

**Physiology measures as an Indicator of Affect**

**Consent Form**

HREC File Number:

Centre: SCHIL, Swinburne University of Technology  
Investigator: Todd Bentley (tbentley@it.swin.edu.au)  
Supervisory Investigator: Lorraine Johnston (ljohnston@it.swin.edu.au)

I __________________________________ have read and understood the information presented about these procedures and risks involved in this study, and have received satisfactory answers to my questions related to this study.

I agree to participate in this activity, realising that my anonymity will be maintained and that I may withdraw from this study at any time. I agree that the activity may be recorded on video tape as data on the condition that no part of it is included in any presentation or public display, without my prior consent.

This study has been reviewed by, and received clearance through, the Human Research Ethics Committee at Swinburne University of Technology. If you have any concerns regarding the ethical nature resulting from your participation in this study, please contact the Chair of the Human Research Ethics Committee:

The Chair  
Human Research Ethics Committee  
Swinburne University of Technology  
P O Box 218  
HAWTHORN. VIC. 3122  
Phone: (03) 9214 5223

I agree that research data collected for the study may be published or provided to other researchers on the condition that anonymity is preserved and that I cannot be identified.

Name of Participant

Signature _________________________________ Date: dd/mm/yy

Investigator(s) Present

Signature _________________________________ Date: dd/mm/yy

Signature _________________________________ Date: dd/mm/yy
E.3 Participant Video Release Forms

**Physiology measures as an Indicator of Affect**  
**Video Release Form**

---

**HREC File Number:**
Centre: SCHIL, Swinburne University of Technology  
Investigator: Todd Bentley (tbtentley@it.swin.edu.au)  
Supervisory Investigator: Lorraine Johnston (ljohnston@it.swin.edu.au)

---

**Information:** This area of study and measurement protocol is new for the field of human-computer interaction. As a result, many practitioners in this field would have limited knowledge regarding physiological measures and the overall setup of this equipment.

For this reason we are seeking your permission to use short video clips from your session to be used for demonstrative purposes during ensuing presentations or publications. In no way will these videos be released, and will remain in possession of the investigators only. Only audiences at these presentations will have the opportunity to observe video clips. No clips will be shown that are potentially unflattering or embarrassing to you. Further, no information will be divulged about results related to your session.

All participants completing this study have been or will be asked for permission to use their video clips, and there is no guarantee that your video clips will be used. You are under no obligation to allow us to use your video clips, and it will not impair your status as a participant in this study.

---

**Consent:** I ____________________________ would be happy for the investigators of the study Physiology Measures as an Indicator of Affect to use small excerpts of video from my session for educational and/or demonstrative purposes. No excerpts would be shown that could be construed as unflattering or embarrassing for me.

Name of Participant

---

Signature ____________________________ Date: __dd/mm/yy__
Investigator(s) Present

Signature ____________________________ Date: __dd/mm/yy__
Signature ____________________________ Date: __dd/mm/yy__
Participant ID#: _____

The purpose of this questionnaire is to obtain general demographic information to allow for more detailed analysis and understanding of study results. Should you have any questions regarding the demographic information collected feel free to ask.

1. **Age Range.**
   - [ ] 18-25
   - [ ] 26-30
   - [ ] 31-35
   - [ ] 36-45
   - [ ] 46-55
   - [ ] 56-65
   - [ ] 66+

2. **Gender.**
   - [ ] Male
   - [ ] Female

3. **Main Profession/Occupation.**
   - [ ] Student
   - [ ] IT (Industry)
   - [ ] IT (Research)
   - [ ] Arts
   - [ ] Legal
   - [ ] Business
   - [ ] Education
   - [ ] Customer Service
   - [ ] Other _____________

4. **How often do you use a computer?**
   - [ ] <1 hour/week
   - [ ] 2-5 hours / week
   - [ ] 6-10 hours/week
   - [ ] More than 10 hours/week

5. **How often do you use a computer for leisure?** (i.e. games, surfing the internet, etc.)
   - [ ] <1 hour/week
   - [ ] 2-5 hours / week
   - [ ] 6-10 hours/week
   - [ ] More than 10 hours/week
6. Where do you usually use a computer for leisure? (select all that apply)
   - Home
   - Work/School
   - Internet Cafe
   - Other

7. What sort of computer games do you play? (select all that apply)
   - Driving/Flying Simulators
   - Role-Playing
   - Strategy
   - Puzzle Games (i.e. tetris)
   - First Person
   - Strategy
   - General Simulation (i.e. the SIMS)
   - Solitaire
   - Other (list all other types)
     __________________________
     __________________________
### E.5 Participant Demographics

**Table D.1: Complete Participant Demographics for Physiology Study**

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Gender</th>
<th>Occupation</th>
<th>Amount of Computer Use (hours/week)</th>
<th>Amount of Computer Use for Leisure (hours/week)</th>
<th>Where use computer for leisure</th>
<th>Preferred Game Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18-25</td>
<td>Male</td>
<td>Student</td>
<td>&gt;10</td>
<td>&gt;10</td>
<td>Home</td>
<td>Driving, FPS, Strategy puzzle, rpg, fps, solitaire, strategy</td>
</tr>
<tr>
<td>2</td>
<td>18-25</td>
<td>Male</td>
<td>Student</td>
<td>&gt;10</td>
<td>&gt;10</td>
<td>Home</td>
<td>puzzle, rpg, fps, solitaire, strategy</td>
</tr>
<tr>
<td>3</td>
<td>18-25</td>
<td>Male</td>
<td>Student</td>
<td>&gt;10</td>
<td>2-5</td>
<td>Home</td>
<td>driving, simulation puzzle, simulation</td>
</tr>
<tr>
<td>4</td>
<td>18-25</td>
<td>Female</td>
<td>Student</td>
<td>&gt;10</td>
<td>&gt;10</td>
<td>Home</td>
<td>puzzle, strategy n/a</td>
</tr>
<tr>
<td>5</td>
<td>26-30</td>
<td>Male</td>
<td>Chef</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>Home</td>
<td>fps, strategy</td>
</tr>
<tr>
<td>6</td>
<td>18-25</td>
<td>Female</td>
<td>Student</td>
<td>&gt;10</td>
<td>&gt;10</td>
<td>Home</td>
<td>simulation, puzzle, simulation</td>
</tr>
<tr>
<td>7</td>
<td>18-25</td>
<td>Male</td>
<td>Student</td>
<td>&gt;10</td>
<td>&lt;1</td>
<td>Home</td>
<td>fps</td>
</tr>
<tr>
<td>8</td>
<td>18-25</td>
<td>Male</td>
<td>IT industry</td>
<td>&gt;10</td>
<td>6-10</td>
<td>Home</td>
<td>fps, strategy</td>
</tr>
<tr>
<td>9</td>
<td>18-25</td>
<td>Female</td>
<td>Student</td>
<td>&gt;10</td>
<td>2-5</td>
<td>Home and work</td>
<td>n/a</td>
</tr>
<tr>
<td>10</td>
<td>26-30</td>
<td>Male</td>
<td>Education</td>
<td>&gt;10</td>
<td>&gt;10</td>
<td>Home</td>
<td>driving, rpg, fps, sport, strategy</td>
</tr>
</tbody>
</table>
E.6  Inter-rater Reliability Forms

The inter-rater reliability forms were provided to three independent raters. These forms are reproduced here.

I would like to thank you for offering to assist me in rating the comments on the attached pages. This will be a tremendous help to my studies. The goal of your ratings is to try and determine if the affective ratings that I have placed on comments made by participants is consistent with the views of other independent observers. This is important to know as this will ensure that the results from the study I performed are not simply artefacts created by my rating scale, rather that they are representative of the comments made by participants. I believe the ratings to be accurate, but a second opinion will add substance to this assertion (assuming, of course, that the ratings end up being consistent).

Thank-you for your help,

Todd

Instructions for rating:

The following pages contain a total of 180 comments made by participants during the study. These are not all the comments collected as some have been removed previously for various analytical purposes. I would like you to indicate whether each comment is representative of a ‘neutral’, ‘positive’, or ‘negative’ affective response. If you do not believe it is part of any of these, feel free to select the ‘generic or non-applicable’ column. These are formally described as:
5. **Neutral**: debriefing comments where the participant indicated that they were not feeling or thinking anything (i.e. "I wasn’t really thinking too much, I was pretty much just going along.");
6. **Positive**: debriefing comments where the participant indicated that they had a feeling or affective response at that time that you believe is positive (i.e. "I really found that part exciting when.");
7. **Negative**: debriefing comments where the participant indicated that they had a feeling or affective response at that time that you believe is negative (i.e. "I was annoyed right there when.");
8. **Generic or Non-Applicable**: debriefing comments that do not fit into any of the above categories. This should only be chosen when there is no reference to any sort of affective responses, or lack thereof. Most times this is chosen when the comment is a description of their overall emotion rather than their short-term affective state. (i.e. "I found the game to be enjoyable")

An affective response is defined as “a short term emotional change, which may or may not have long term impacts on a person’s overall emotion”. Affect needs to be distinguished from an emotion which is the “person’s overall feeling and has a variety of influences acting upon it” – one of which can be affect.

It is my experience that there are some affective responses that could, depending on how you look at them, be classified as either positive or negative. In these instances I would urge you to select the one you feel is best – DO NOT select generic in this situation. If you feel you need to explain your decision please make note of the comment number, and provide comments either on the back of the pages or if there is room you may write on the front of the pages.

An example rating of a comment is shown. You can feel free to select the appropriate column in any manner you desire. X’s, check marks, highlights are all acceptable. The only requirement is that I will be able to discern your choice afterwords.

<table>
<thead>
<tr>
<th>Ex.</th>
<th>Context of Comment</th>
<th>Debrief Comment</th>
<th>Neutral (no affect)</th>
<th>Positive affect</th>
<th>Negative affect</th>
<th>Generic Comment or Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetris: Getting near top</td>
<td>That really irritated me. I was waiting for a block but it just wasn't coming.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
I have provided a context of comment to try and assist your decision making. There are times where the comments cannot be fully understood without having some idea of the context within which it is made. The context also includes the name of the game that the participant was commenting on. This is either Tetris or NFS (a Driving / Racing game; NFS = Need for Speed). If there is no context provided after the game name it means that nothing was remarkable at that stage of the game.

Additional Notes:

The comments have been randomized. They are not grouped by participant, game type, or any other possible grouping. I should also mention that I have left the interpretations of comments relatively unrestrictive as I did not want to bias your decisions in any form.

There are some comments that you will be unsure about which column to choose. I can tell you that there were some of these for me too. This especially is true when you know it fits one of the categories, but really have no idea which one it best suits (and then you start thinking ‘maybe none’). My only advice is to simply choose whichever one you think is best and move on. I will also mention that I am more concerned whether there was or was not an affective response (i.e. neutral vs. <positive or negative>). I am asking you to choose positive and negative to allow me to explore the data in further detail.
<table>
<thead>
<tr>
<th></th>
<th>Game (and event if applicable)</th>
<th>Debrief Comment</th>
<th>Neutral (no affect)</th>
<th>Positive affect</th>
<th>Negative affect</th>
<th>Generic Comment or Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tetris</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>NFS: Went wide on a corner</td>
<td>Things just seem to be going smoothly at this point</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Tetris: put many ‘wrong’ blocks down</td>
<td>That didn't really bother me - it was my fault.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tetris: recover from poor situation, got some ‘big blocks’</td>
<td>That's the one - just total frustration. Just totally stuffed it up there.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Tetris: level change</td>
<td>I would have been happy at that point</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Tetris: near end of game.</td>
<td>The level changed there and made me kinda sit up and I was expecting things to be different there.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>NFS: Participant lost control of car and was facing backwards.</td>
<td>This is the point where I just gave up and essentially resigned myself that the game was over.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Tetris: at beginning of game, blocks falling slowly</td>
<td>I found that bit frustrating because whenever that happens I can never manage to turn myself around the right way. I ended up going 100 kilometers in the opposite direction before I can get turned back around.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>NFS: straight portion of track</td>
<td>I did find it a bit too slow at this point, but nothing else. I assumed it would speed up a bit later.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>NFS: hit ‘something’ and lost control</td>
<td>This was an easy-going bit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Tetris</td>
<td>That was annoying. I was approaching other cars and thought I was doing alright, but then something happened.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>NFS: started out and passed all cars</td>
<td>I had a lot of time to think about each piece at this point, and generally didn't get too worried about anything happening as I had lots of time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Game (and event if applicable)</strong></td>
<td><strong>Debrief Comment</strong></td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>13</td>
<td>Tetris: Just started</td>
<td>Wicked, I passed all the cars. It was fun because you could take them on the inside too.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Tetris</td>
<td>The level was a bit easy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Tetris: going smoothly, despite level change</td>
<td>I had just got a few tetrisses a while back which was good, and it made this part of the game a lot easier to play as I was able to plan where pieces would be put.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>NFS: Participant crashed using a shortcut in the game. Didn't think they would hit the wall but they apparently did.</td>
<td>I was going pretty good at that point. Seems like I was starting to get things back lower. and you know, the next level and everything. I was feeling pretty good at this point and thought I was doing well</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>NFS</td>
<td>I was a bit annoyed there</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>NFS: game loading; game loads slowly which is what the comment is referring to</td>
<td>I was able to relax at these points.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>NFS: jumped over rise in road</td>
<td>yeah, I was a bit confused because it didn't do much while it was loading - wasn't sure if it was stuck or something. It could show me my car spinning around or something.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Tetris: participant got a few blocks in a row getting the level down to baseline.</td>
<td>I was surprised when I got some air there - it was a fun! I guess I was sort of expecting it to happen - I played these games before before and seem to remember that I could do that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>NFS: Participant missed shortcut</td>
<td>I like those bits - it feels good when I get it all back neat and orderly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>22</td>
<td>NFS: at the beginning of the game the participant sped ahead to the front of the pack</td>
<td>Yeah, I actually remember that shortcut from the first round. I would have liked to take that, and it was a bit frustrating missing it a second time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Tetris: participant was back down at the bottom</td>
<td>I was definitely feeling confident right there. I went to the front straight away and was feeling good about that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Tetris: near beginning, slow</td>
<td>I was down at the bottom and was thinking 'no worries'.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>NFS: crashed car while passing another</td>
<td>I thought it was a bit easy at that point.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>NFS: game ended. It was not apparent where the stop line was.</td>
<td>While I was approaching that car I was just hoping to hold the line, and then I realised that my car wasn't turning. I then realised that I must have damaged it. It was pretty frustrating because I thought I had a chance there too. Then the car stopped responding, and I wasn't happy with myself.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Tetris: beginning of game, blocks are slow.</td>
<td>That was confusing. I didn't realise I was coming up to the end there. I saw the wall and then the car stopped. It was a bit abrupt. After a bit I realised what had happened.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Tetris: started to get high on board</td>
<td>I wasn't really thinking anything. I was just waiting until the game got exciting.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>NFS: Participant flipped the car</td>
<td>That is when I get irritated and want the game to end and start refreshed. I don't like losing so when it gets that high I know I am close and would rather be down low again, even if that means restarting.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>30</td>
<td>Tetris: beginning of game, blocks are slow.</td>
<td>I thought that was great! I was sort of thinking it might go over the side. It was funny and good.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>NFS</td>
<td>I found the level a bit slow, but when I got to higher levels I crashed and burned. But it isn't too bad. If you wanted to speed them up all you have to do is press space and they will speed up to the level you want them to go.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Tetris: game had just finished.</td>
<td>I was just going along. It wasn't really fun at that point, but not boring either.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>NFS: passed all cars at beginning</td>
<td>It was over, and that meant I could relax now. I was just waiting for the end - you have about a minute or two knowing the end is near, but hope you can get through even though you know you can't.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>NFS: approaching other cars but then another car pushed participant off the road.</td>
<td>I was pretty happy about that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Tetris</td>
<td>Whoops. That was about it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Tetris</td>
<td>Thought this section was pretty straight forward.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Tetris: 4 straights in a row, helping out a lot</td>
<td>Though I was still doing, yeah, calm at this stage. It wasn't even halfway up the screen yet.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>NFS: game loading</td>
<td>I felt much better at that point. I got lucky with a few blocks.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>39</td>
<td>NFS: near end of game. Entire level went smoothly (too easy for participant)</td>
<td>I was just waiting for the track to load, and wondering if experts (re: game playing experts) try to study these tracks when they see them.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>NFS: heading into town with tight turns</td>
<td>This is when I decided this game didn't appeal to me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>NFS: Game loading</td>
<td>This is the most exciting bit of the course. You come around the corner just where the ships are and crash into something. You tend to be going quite fast and it hits you by surprise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>NFS: lost control of car. But still extremely far ahead of everyone else.</td>
<td>This is the bit where I was sort of looking forward to what was coming next, and I wasn't quite sure what to expect. I think it is like the previews at a movie or something.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>NFS: was going well, but then for some reason went straight into wall</td>
<td>This is the bit when I went 'it doesn't matter'.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>NFS</td>
<td>I think that was the bit I was laughing. I crashed there, and that just doesn't happen in real life (re: car response after crashing). Just thought it was funny.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Tetris: game speeding up a lot</td>
<td>This bit I was doing well. Making all the turns and everything going as planned.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Tetris</td>
<td>I felt that things were getting a bit away from me at this point. I couldn't keep up and got a bit agitated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>NFS: Game was loading</td>
<td>Very happy - I got two big tetrisses and that made the next bit of the game so much smoother.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>48</td>
<td>NFS: cars ahead seem to take off faster. Shortly thereafter participant passes all cars though.</td>
<td>I was thinking &quot;I know I am going to suck at this a lot&quot; and really was getting nervous about someone watching me play this game. I am a terrible driver in real life, so playing this style game just compounds it.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>49</td>
<td>Tetris: tapped a few pieces down in a row - recognised mistake</td>
<td>This is the bit when the lead car started to pull away and I thought &quot;Excellent, this is going to be a race&quot;. The last round the cars were just to easy to beat.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>50</td>
<td>Tetris: got a few big blocks. Reduced height a fair amount</td>
<td>I just lost concentration there a bit and pressed space. I don't know why - I just zoned out I think.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>51</td>
<td>NFS: A slight pause in game (glitch) but then kept going smoothly</td>
<td>There I had a few strings of pieces. That was cool.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>52</td>
<td>NFS: big wipe out, quick recovery</td>
<td>The computer seemed to have a glitch there, which was strange because it was the only one I had seen.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>53</td>
<td>Tetris: getting higher on board</td>
<td>That felt good when I recovered like that. Made you think you know what you are doing.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>54</td>
<td>NFS: Crashed car</td>
<td>That was when I was stressed because things were building up and I wasn't getting anything that I wanted.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>55</td>
<td>NFS: smooth section of track</td>
<td>That was the bit where I stacked it - that was really fun.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>56</td>
<td>NFS: hit wall (this wall was 'hidden' from the point of view of participant in game)</td>
<td>That felt good. I was thinking &quot;I hope you (experimenter) are watching this, this is good&quot;.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>57</td>
<td>NFS: at beginning, approached cars, passed them all.</td>
<td>That was frustrating, things like that. I had no warning that those walls were coming up. I didn't know what happened at the time, but now I realise I hit a dividing wall or something.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>58</td>
<td>NFS: crashed shortly after passing cars.</td>
<td>That was cool. I remember hitting the back of some of the cars, and was able to just keep going through.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>NFS: passed all cars at beginning</td>
<td>That was a frustrating crash right there. Some don't bother me because I can keep going, but that one did for some reason. I think it was because it let all those cars back around me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>NFS: track was easy, and then suddenly became difficult</td>
<td>I thought it was going to be a bit easy at that point.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>Tetris: level change</td>
<td>(last comment: I thought it was going to be a bit easy at that point). Until I came to turns like this and realised it was going to a bit more exciting with tricky turns like this.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>Tetris: participant placed a block poorly.</td>
<td>I was pretty happy that I reached the next stage.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>NFS: participant hits another car and the participant slowly loses complete control of the car.</td>
<td>That was it! That is where I put the wrong block down, and really bugged the rest of the game. Annoyed right there.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>Tetris: near end of game.</td>
<td>That was a bit annoying.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>NFS: crashed at beginning of first round</td>
<td>At this point I was just thinking &quot;one more line&quot; and that was about it. I realised I was going to lose soon, but just kept trying for that last line.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>NFS: the participant crashed just before finish line. The participant then drove forward a little, and the game 'froze' (but really participant had just crossed finish line)</td>
<td>That bothered me a bit. I was still trying to figure out the controls but they didn't seem to be responding properly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>67</td>
<td>Tetris</td>
<td>I had no clue what was happening there. I was totally confused.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>NFS: coming up towards end, but not making headway on car in front.</td>
<td>Happiness there - that is the play that brought me back down.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>Tetris: start of game</td>
<td>This is where I was thinking &quot;crap, I'm going to lose&quot;.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>NFS: start of game</td>
<td>Nothing really - just trying to place the blocks.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>Tetris</td>
<td>I was looking forward to playing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>NFS: game loading (second round)</td>
<td>I thought I can usually play tetris quite a bit faster so I wasn't worried or anything at this point.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>NFS: game ended. It was not apparent where the stop line was.</td>
<td>I wasn't anticipating this round like I was last time. I knew what to expect this time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>NFS</td>
<td>It was just the ending. That is about it - just the end of the game.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>NFS</td>
<td>Not really thinking anything actually.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>Tetris</td>
<td>Things were going smoothly, and no real surprises.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>NFS: car flipped</td>
<td>By this stage, it was pretty apparent that I wasn't going to get back down to the lower levels and started to get a bit more frustrated. you can see it slowly building up.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>Tetris</td>
<td>This was good - it was working out again.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>NFS</td>
<td>That was good - it was working out again.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>NFS: approaching cars, not steering around them</td>
<td>I was sort of going with the flow.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>81</td>
<td>NFS: game loading</td>
<td>Moment of panic - &quot;quick, get out of my way&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>NFS</td>
<td>There was some anticipation when the game was loading. It took so long it seemed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>NFS</td>
<td>There was a little shortcut there. I thought I had understeered there and was ready to go off the road but then I saw the shortcut and all was fine.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>NFS</td>
<td>It was just going really smooth at this point. I think I lost it up ahead, but it was smooth at this point.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>NFS</td>
<td>Just playing really.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>Tetris</td>
<td>Just cruising along.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>NFS</td>
<td>I was just mentally comparing it with other tetriss I had played before, that was about it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>Tetris; starting to build up slightly</td>
<td>It wa all part of the game pretty much. It was perhaps relaxing - gave me time to think ahead.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>NFS</td>
<td>It's all part of the game. No surprises yet.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>Tetris</td>
<td>You are in some sort of head space where everything is going really well. It is kind of an easy section of the track I think too. Big turns and stuff. I was able to just go along.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>NFS; for the first time got into first place</td>
<td>It was a bit slow, but I didn't need much time so I just used the space bar a lot to speed it up.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>Tetris</td>
<td>It was really rewarding at this point.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>93</td>
<td>NFS</td>
<td>It was getting faster around here. I was getting quite nervous at this point I think. The right bricks weren't falling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>Tetris: near top of board</td>
<td>I had just passed that car, and saw another up ahead. I thought it would be a good challenge to see if I could catch them.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>Tetris</td>
<td>It was a little stressful.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>Tetris: pressed space at an incorrect time</td>
<td>It took me a while to get into it. But here I had just gotten a few good blocks, and got a bit excited because I felt like I was getting into the groove of it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>NFS</td>
<td>It stressed me out a bit when I hit that button incorrectly. I thought it was a line, but didn't line it up well enough and got frustrated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>Tetris: level change</td>
<td>I didn't know which way to go on that fork in the road. I decided to stay on the main road simply because it was just a continuation of the road. It didn't bother me, just didn't know what to do fully.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>NFS: Game loading</td>
<td>That made me sit up a bit and made me think 'ahh, doing well'. I was expecting an immeditated change in the game - for it to be suddenly quicker and pieces flying down a lot quicker.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>NFS: crashed into a wall</td>
<td>It just felt like it was taking a while and was worried that I may have crashed the computer.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>Tetris</td>
<td>It didn't bother me when I was crashing. The idea is that you play it a few times, and learn where the turns are, and the traffic islands and stuff like that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>NFS</td>
<td>I found things pretty slow here. I was trying to get used to things again.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>103 NFS</td>
<td>I just seemed to be going along. A bit of luck here and there maybe.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104 NFS: end of game. No 'finish line' on this track</td>
<td>I was getting close to passing him, but I was also close to the end. I didn't want to smack into him because I would just take myself out as well. It was sort of nerve-racking - do I pass him now or later.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105 Tetris</td>
<td>I wasn't too sure it was the end. I thought I might have had another round or something, and then I just suddenly stopped. Confused me quite a bit until I figured out what happened.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106 NFS</td>
<td>I wasn't thinking a great deal. Just trying to find places to put them in place.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107 NFS</td>
<td>I wasn't really thinking much during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>108 Tetris</td>
<td>I wasn't really feeling much around here. I was pretty comfortable with everything at this point and was just going along. Happy to have gotten into the groove of things, and pretty comfortable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109 Tetris</td>
<td>I wasn't getting too nervous because there were not too many gaps. I always do slow it down though when I get near the top and play a bit more careful. I pay attention to the next piece more and determine my next move as well</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110 NFS: constantly switching positions with another car until end of round</td>
<td>I wasn't getting piece 2 I needed though and that made it frustrating. I am sure that the computer is programmed to give you every piece except for the ideal piece.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>111</td>
<td>Tetris</td>
<td>I was thinking that it was getting more fun because it was just the two of us and we were just racing. It was manageable too - it wasn't as though he passed me and he disappeared into far away. So that was more fun and it was nice beating it, and still having it there to challenge me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>Tetris</td>
<td>I was quite confident and I was just thinking that the means justifies the ends. Meaning that if I left a gap it doesn't matter as long as I get a few more rows. So far I was feeling pretty good with my game.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>Tetris: level change</td>
<td>I was pretty annoyed there when I put the block down in the wrong spot. It always happens once or twice in tetris though.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>NFS: smooth portion of track, and approaching end</td>
<td>I was happy there. Because I was on level 4 and going strong. Close to the bottom.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Tetris</td>
<td>I was getting all excited there!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>NFS</td>
<td>I was feeling a bit embarrassed more than anything. Because of no real gkr gkpegy kj <code>guk/cpf </code>cmA `ngr k cwp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>NFS</td>
<td>Everything was just going smoothly and finely.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>NFS</td>
<td>I was a little bit excited.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>NFS</td>
<td>I was thinking that the game was not that hard at all.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>NFS: game just started, and the other cars took off faster.</td>
<td>That was funny. I just headed straight for that tree. I thought I was trying to avoid it, but it just didn't work!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>NFS: participant chose a vehicle that allowed him to see face of driver</td>
<td>The other cars took off a bit better, and just seemed better generally. I was driving ok, so I thought I should have picked cpq g&quot;ect0&quot;DswY cvf k plv/dqy g&quot;o g&quot;. I just went along looking at the scenery in the game.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>122 NFS</td>
<td>When I saw the character in the game I found it funny. His face was just a bit too square.</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>123 Tetris</td>
<td>Kj qw j vK cf &quot;c1 qqf &quot;ej cpeq0kK k plhj lphj g&quot;ect&quot;y cu\qq&quot; bad. I overtook them all right away which made me feel good.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>124 NFS</td>
<td>I think this is the first time I realised I was going to die and was panicking. I was in trouble. I was just trying to find a place to place each block. Rushing my thinking</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>125 Tetris: level change</td>
<td>This was where I decided to try a bit and play seriously - stop mucking around.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>126 Tetris: participant was at the bottom of the board again</td>
<td>I thought that this was getting a bit tougher here - kept speeding up. I was getting worried because it started going a bit too fast.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>127 Tetris</td>
<td>I think I was probably just relaxed I would say. I wasn't too y qtlqf &quot;cdqwj qy &quot;Kj cu\qi qlo \qi i&quot;.\uqA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>128 NFS</td>
<td>I think I was pretty much just going along at this point.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>129 Tetris</td>
<td>I think I was just more confident this round. Wasn't worried because I knew what I needed to do and what to expect.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>130 Tetris: participant was waiting for a block for about 1 minute, did not come.</td>
<td>I think I was in the zone there.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>131 NFS: participant saw and passed lead car</td>
<td>I got frustrated around here, because the block that I wanted wasn't coming down - the long straight one - and I was panicking a bit.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>132 Tetris: level change</td>
<td>I saw him quite a bit before, but then lost him. But then he came up again and I really wanted to get in front. When I did it was pretty fun.</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>133 Tetris: participant was relying on block, not coming, so gave up on it.</td>
<td>I remember being annoyed at the change to level 5. It happened at a really bad time because I was about half way up and the pieces really sped up.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>134 NFS</td>
<td>I kept saying &quot;when is a long one going to come along?&quot; And eventually I just decided to just say 'screw it' and just went ahead. It bothered me that I had to do that. And then two blocks later it came along. Should've waited.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135 NFS</td>
<td>I got lost there. Had no idea where to go so I ended up just following the other car.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>136 Tetris: several gaps were left</td>
<td>I found this to be a pretty boring part of the track. I remember listening to, you can hear all the sounds here and I remember listening to them. I was just going I guess.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>137 Tetris: at the bottom of the board</td>
<td>I found myself starting to get frustrated and more agitated because I knew I was going to die soon.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>138 NFS</td>
<td>I find I relax when it gets down like that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>139 Tetris: left first gap</td>
<td>I ended up losing control right there, and I knew I was. That's when I went 'o damn'.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140 Tetris</td>
<td>I don't think it really bothered me that much.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>141 Tetris: level change</td>
<td>I was still in full control at this stage.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>142 Tetris</td>
<td>The level changed just a while back and I was initially expecting a big change. But at this point I was thinking &quot;that's not that big of a change, I can handle this&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>143</td>
<td>Tetris</td>
<td>I didn't really know what I was thinking then. I guess I was just trying to come up with a strategy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>NFS</td>
<td>I wasn't worried when I had built up to halfway. Only when it gets higher does it become a problem. I didn't have any gaps either, so it was fine.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>NFS</td>
<td>Here I took the wrong turn. On the map on the loading screen I saw a shortcut, and realised I took the long way around. Then I looked at the map on my screen and realised that one of the other cars caught up to me by taking the shortcut which really annoyed me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>146</td>
<td>NFS: flipped car</td>
<td>Here I am just cruising along, thinking &quot;yeah, I am just cruising along, doing pretty good&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>NFS</td>
<td>Ha ha. That's where I rolled the car.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>148</td>
<td>NFS:</td>
<td>Got some air on that jump which was cool.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>NFS</td>
<td>Found this car much easier to control, but nothing really other than that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>NFS</td>
<td>I was just in a zone there.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>NFS</td>
<td>That caught me by surprise! I didn't see that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>Tetris: getting higher on board</td>
<td>I thought 'how embarrassing to crash straight into a wall'. I didn't even think to move around it and was embarrassed that you had to see that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>NFS</td>
<td>There was a bit of adrenalin there. I could feel it. I was concentrating more, and more aware.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and ‘event’ if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>154</td>
<td>Tetris: almost at top</td>
<td>That was interesting, that jump. Didn’t expect that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>NFS: game loading</td>
<td>I was ready to just give up and start over. I was annoyed because there were too many gaps and thought it would just be easier to start over.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>156</td>
<td>Tetris</td>
<td>At that point I was going ‘cmon’. It was slow to load, and I was looking forward to it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>157</td>
<td>NFS</td>
<td>Around here I was thinking I would be in a bit of trouble. Probably a bit stressed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>158</td>
<td>NFS:</td>
<td>And then we hit town and this is where I think ‘crap - this is usually where I hit the walls’. And sure enough I do. How annoying is that - I know where I screw up but I still screw up.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>159</td>
<td>NFS: lost control, but hit other cars as well</td>
<td>And then that car passed me which was really annoying. Throughout the whole game none of them were close.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>NFS: second place car passed participant right near end of race</td>
<td>I spun out of control and was sort of annoyed again. But it was sort of funny too because I took out some other cars so I still stayed in second place.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>161</td>
<td>NFS</td>
<td>I was definitely not happy there. He just passed me and I thought I wasn’t going to catch him again. So close, yet so far.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>162</td>
<td>Tetris: not getting the block needed</td>
<td>I got stressed again after that.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>163</td>
<td>NFS</td>
<td>And that is where things go wrong. I was waiting for that bloody 60°Cpi 'KwuNgr' 1 qdi 'Y co pÂ OF co pÂ 0Fq'gej 'dmem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>----------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>164</td>
<td>NFS: following a car for the past while</td>
<td>Nice smooth portion of the track. Nothing too exciting happened here.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>NFS: game loading</td>
<td>And here I thought &quot;here's my chance to pass him&quot; and got a bit excited when I did.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>NFS</td>
<td>And especially while the loading screen was coming up, it was a bit slow, I was feeling with anticipation to get into it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>NFS</td>
<td>And around here I was thinking that the car was really slow, and it makes it kind of a V-W. Oh-well, just enjoyed the track and graphics.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>168</td>
<td>NFS</td>
<td>Nothing (response to prompt)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>NFS: crashed into a wall</td>
<td>Ah - I was doing so good in this game! (referring to the fact that she was in a very smooth portion of the game, no accidents, passing many cars, etc)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>Tetris: level change</td>
<td>A little frustrating, but more funny than anything else. Because it is just a game, so I didn't take too much notice of it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>NFS: near beginning of round one</td>
<td>C&quot;łu</td>
<td>[łkA &quot;gzeki</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172</td>
<td>NFS</td>
<td>A bit confused there.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>173</td>
<td>NFS: flipped car</td>
<td>This is bothersome- I always seem to get stuck like this (comment made during game-play session, not debrief)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174</td>
<td>Tetris</td>
<td>I just thought &quot;huh?&quot;, and tried to figure out what happened. I know I flipped, but had no idea why.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Game (and event if applicable)</td>
<td>Debrief Comment</td>
<td>Neutral (no affect)</td>
<td>Positive affect</td>
<td>Negative affect</td>
<td>Generic Comment or Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>---------------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>175</td>
<td>NFS</td>
<td>That bothered me for about two seconds.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>176</td>
<td>Tetris: starting to build up</td>
<td>I was starting to get comfortable with the controls a bit more. Starting to realise that I have to slow down to take corners. This section ended up going really well because I just took my time and didn't try to push it and end up making mistakes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>177</td>
<td>NFS</td>
<td>Things started to build up a bit, but I thought &quot;I can still handle this&quot; and just kept playing, not really concerned.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>Tetris</td>
<td>I was going really well here. Strange - when I go really well I think great, but I only really get excited at the tough bits or when I have an immediate challenge. But yet I always look forward to the smooth bits when I can just relax and not think. At the same time, the game would end up being boring if it was all smooth.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>179</td>
<td>Tetris</td>
<td>Got things back down, and things were going well.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>NFS</td>
<td>Woo hoo! I got the block I needed (badly). That fixed my last mistake.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## E.7 Inter-Rater Reliability Assessments

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>26</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>44</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>47</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Rater 1</td>
<td>Rater 2</td>
<td>Rater 3</td>
<td>Original</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>53</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>54</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>56</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>57</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>58</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>59</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>61</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>62</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>63</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>64</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>65</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>66</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>67</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>69</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>71</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>73</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>74</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>76</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>77</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>78</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>79</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>80</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>81</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>82</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>83</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>86</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>90</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>92</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>93</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>94</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>95</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>96</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>97</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Rater 1</td>
<td>Rater 2</td>
<td>Rater 3</td>
<td>Original</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>98</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>99</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>101</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>104</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>105</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>106</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>107</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>108</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>109</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>110</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>111</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>112</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>113</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>114</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>115</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>116</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>117</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>118</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>119</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>120</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>121</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>122</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>123</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>124</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>125</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>126</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>127</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>128</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>129</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>130</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>131</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>132</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>133</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>134</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>135</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>136</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>137</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>138</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>139</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>140</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>141</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>142</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>143</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>144</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>145</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>146</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
### E.8 Inter-Rater Reliability Results

Inter-rater reliability analysis for the classification of affective comments.

1. RATER1
2. RATER2
3. RATER3
4. RATER0

Reliability Coefficients

N of Cases = 175.0
N of Items = 4

Alpha = 0.9831
List of Publications during Thesis Process

The following works were completed during the development of this thesis. Several of these were researched and written in relation to possible thesis paths, however based on the results of the research or due to the refinement of thesis goals they were no longer related to the major goals presented within this thesis.


