

**A behavioural analysis of computer game playing competence,
experience and related physiological processes**



NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Thesis submitted to the Department of Psychology, Faculty of Science, National University of
Ireland, Maynooth in fulfillment of the requirements for the degree of

Doctor of Philosophy (Psychology)

April 2008

Conor Linehan, B.A.

Supervisor: Dr. Bryan Roche

Acknowledgements

Thanks a million to the following...

Bryan, this was your idea in the first place, I'm blaming you! hahahaha! Seriously though, for the ideas, the expertise, the enthusiasm and most importantly the huge amount of time you've put into this, I cant thank you enough. Without your help I simply wouldnt be handing this in.

Eamonn, for the technical wizadry and sheer volume of time put into the formatting and style of the thesis.

Tomás Ward for hooking me up with this project and taking the time to consider the ideas of lowly psychologists

The Irish Research Council for Science, Engineering and Technology/Embark Initiative for the cold, hard cash that lured me back to academia.

The poor souls who took the time to proof read a chapter or a section for me: Christine, Paraic, John and Keo. Pedanticism at its most useful!

The poorer souls who had to sit through any of these experiments! This would not have been possible without you.

The occupants of 4 Castlebridge and the rest of my rag-tag bunch of friends and hangers-on for de-stressing value.

And of course, to my Mum and Dad for your love and support throughout.

Conor Linehan

National University of Ireland, Maynooth

April 2008

Abstract

The current research programme represents a first step in the psychological analysis of on-line game playing. In the literature review presented in Chapter 1, Network Latency and 'game challenge' were identified as two important variables affecting participants' enjoyment of on-line games. The experiments presented in Chapter 2 defined 'game challenge' in terms of levels of derived relational responding, and found that participants were able to consistently respond in accordance with both one and three-node derived relations in the context of a computer game. The presence of Network Latency in a game was found to be detrimental to the game playing experience, but increasing the length of those delays was not. The experiments presented in Chapter 3 defined 'game challenge' in terms of more complex forms of derived relational responding and found that participants were able to consistently respond in accordance with derived 'Same' and 'Opposite' relations in the context of a computer game. As in Chapter 2, the presence of Network Latency in a game was found to be detrimental to the game playing experience, but increasing the length of those delays was not. Participants were more successful at and preferred the simpler levels of the games examined in Chapter 3. Experiments in both Chapters 2 and 3 successfully modeled on-line game playing in terms of derived relational responding. The experiments reported in Chapter 4 were conducted in order to develop novel behavioural and physiological measures of enjoyment in game playing. It was found that participants' preference for games of varying difficulty was dependent on their experience with those games. In addition, a novel methodology was developed for analyzing electro-dermal activity, which successfully differentiated games on the basis of the preference shown for them by participants. Finally, Chapter 5 reviewed the relevance of the research findings to the research literature.

Contents

Acknowledgements	ii
Abstract	ii
List of Tables	ix
List of Figures	xiii
Chapter 1 Introduction	1
1.1 What is a game?	2
1.1.1 Are Computer games different from traditional games?	3
1.1.2 What makes a game a good game?	3
1.1.3 Challenge and Problem-solving	5
1.2 Approaches to measuring enjoyment in computer games	7
1.2.1 Questionnaires	7
1.2.2 Physiological recording	8
1.2.3 Behavioural Measures	9
1.2.4 Issues specific to Online Games	10
1.3 The Beginnings of a Psychology of Gaming	13
1.3.1 The Science of Behaviour Analysis	14
1.3.2 A Derived Relations Approach to Complex Behaviour	20
1.3.3 Relational Frame Theory: Extending the Derived Relations approach to Complex Behaviour	23
1.3.4 How are derived relations relevant to computer gaming?	24

Chapter 2 Experiments 2.1-2.4	27
2.1 Experiment 2.1	27
2.1.1 Method	33
2.1.1.1 Participants	33
2.1.1.2 Materials	33
2.1.1.3 Design	33
2.1.1.4 Procedure	33
2.1.2 Results and Discussion	40
2.1.2.1 Stimulus Equivalence Training and Testing	40
2.1.2.2 Game Playing	40
2.2 Experiment 2.2	45
2.2.1 Method	46
2.2.1.1 Participants	46
2.2.1.2 Materials	46
2.2.1.3 Design	46
2.2.1.4 Procedure	46
2.2.2 Results and Discussion	47
2.2.2.1 Stimulus Equivalence Training and Testing	47
2.2.2.2 Game playing	47
2.3 Experiment 2.3	51
2.3.1 Method	55
2.3.1.1 Participants	55
2.3.1.2 Materials	55
2.3.1.3 Design	56
2.3.1.4 Procedure	56
2.3.2 Results and Discussion	60
2.3.2.1 Training level	61
2.3.2.2 Short Delay Game	61
2.3.2.3 Long Delay Game	62
2.4 Experiment 2.4	66
2.4.1 Method	67
2.4.1.1 Participants	67
2.4.1.2 Materials	67

2.4.1.3	Design	67
2.4.1.4	Procedure	68
2.4.2	Results and Discussion	68
2.4.2.1	Training level	69
2.4.2.2	Short Delay Game	69
2.4.2.3	Long Delay Game	69
2.5	General Discussion	73
Chapter 3 Experiments 3.1 and 3.2		81
3.1	Experiment 3.1	81
3.1.1	Method	92
3.1.1.1	Participants	92
3.1.1.2	Materials	92
3.1.1.3	Design	92
3.1.1.4	Procedure	92
3.1.2	Results and Discussion	100
3.1.2.1	Relational Pre-Training	100
3.1.2.2	Relational Training and Testing	100
3.1.2.3	Gaming	102
3.2	Experiment 3.2	114
3.2.1	Method	115
3.2.1.1	Participants	115
3.2.1.2	Materials	115
3.2.1.3	Design	115
3.2.1.4	Procedure	116
3.2.2	Results and Discussion	118
3.2.2.1	Relational Pre-Training	118
3.2.2.2	Relational Training	118
3.2.2.3	Gaming	119
3.3	General Discussion	129
Chapter 4 Experiments 4.1-4.3		134
4.1	Experiment 4.1	134
4.1.1	Method	142

4.1.1.1	Participants	142
4.1.1.2	Materials	142
4.1.1.3	Design	143
4.1.1.4	Procedure	143
4.1.2	Results and Discussion	148
4.1.2.1	Main effect: Speed of Presentation	162
4.1.2.2	Main Effect: Stimulus Number	163
4.2	Experiment 4.2	166
4.2.1	Method	168
4.2.1.1	Participants	168
4.2.1.2	Materials	168
4.2.1.3	Design	168
4.2.1.4	Procedure	169
4.2.2	Results and Discussion	173
4.3	Experiment 4.3	180
4.3.1	Method	181
4.3.1.1	Participants	181
4.3.1.2	Materials	181
4.3.1.3	Design	181
4.3.1.4	Procedure	181
4.3.2	Results and Discussion	182
4.3.3	General Discussion	194
Chapter 5 General Discussion		204
5.1	Chapter 2 - Experiments 2.1 and 2.2	204
5.1.1	Network Latency	205
5.1.2	Relational Complexity	205
5.2	Experiments 2.3 and 2.4	206
5.2.1	Relational Complexity	207
5.2.2	Network Latency	208
5.3	Chapter 3 - Experiments 3.1 and 3.2	209
5.3.1	Experiment 3.1 - Relational Complexity	209
5.3.2	Experiment 3.2 - Network Latency	212
5.3.3	General Comments on Chapters 2 and 3	213

5.4	Chapter 4 - Experiments 4.1-4.3	213
5.4.1	Experiment 4.1	214
5.4.2	Experiment 4.2	215
5.4.3	Experiment 4.3	216
5.4.4	General Comments on Chapter 4	216
5.5	Concluding Comments	218
Appendix A The Day Reconstruction Method (DRM)		236
A.1	Instrument Documentation	236
A.1.1	Instruments	237
A.1.2	Administration	237
A.1.3	References	238
A.2	Packet 3	239
Appendix B Experiment 7 Single Subject SCL Analysis		242
Appendix C Sequence of games played by each participant in Experiment 4.2		253

List of Tables

2.1	Number of attempts required by each participant to pass each stage of Stimulus Equivalence Training and Testing	40
2.2	Mean scores for all measures employed on the Training Level of both the Non-Delay (n=10) and Delay (n=10) Games.	41
2.3	Number of Correct Responses made by each participant in the Generalisation Level	42
2.4	Mean scores for all measures employed on the Generalisation Level of both the Non-Delay (n=10) and Delay (n=10) Games.	42
2.5	Number of Correct Responses made by each participant in the Equivalence Level	42
2.6	Mean scores for all measures employed on the Equivalence Level of both the Non-Delay (n=10) and Delay (n=10) Games.	43
2.7	Mean scores for all measures employed during all three levels of both the Non-Delay (n=10) and Delay (n=10) Games.	43
2.8	Results of a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Correct Responses, difficulty, enjoyment and frustration (n=20).	44
2.9	Number of exposures required by all participants to pass each stage of Stimulus Equivalence Training and Testing phase	48
2.10	Mean scores for all measures employed on the Training Level of both the Non-Delay (n=10) and Delay (n=10) Games.	49
2.11	Number of Correct Responses made by each participant in the Generalisation Level	49
2.12	Mean scores for all measures employed on the Generalisation Level of both the Non-Delay (n=10) and Delay (n=10) Games.	50
2.13	Number of Correct Responses made by each participant in the Equivalence Level	50

2.14	Mean scores for all measures employed on the Equivalence Level of both the Non-Delay (n=10) and Delay (n=10) Games.	51
2.15	Mean scores for all measures employed during all three levels of both the Non-Delay (n=10) and Delay (n=10) Games.	51
2.16	Results of a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Correct Responses, difficulty, enjoyment and frustration (n=20)	52
2.17	Number of exposures required to pass each stage of Stimulus Equivalence Training and Testing	60
2.18	Mean scores for all measures employed on the Training Level of both the Short Delay (n=11) and Long Delay (n=11) Games.	61
2.19	Correct Responses made by each participant across the one-node and three-node Levels in the short Delay Game	62
2.20	Correct Responses made by each participant across the one-node and three-node levels in the long Delay Game	62
2.21	Means for all measures employed across the Training, One-Node and Three-Node Levels, across the Short Delay (n=11) and Long Delay (n=11) Games.	63
2.22	Results from a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=22)	65
2.23	Number of exposures required to pass each stage of Stimulus Equivalence Training	68
2.24	Mean scores for all measures employed on the Training Level of both the Short Delay (n=11) and Long Delay (n=11) Games.	69
2.25	Total Correct Responses made by each participant across the One-Node and Three-Node Levels in the Short Delay Game	69
2.26	Total Correct Responses made by each participant across the One-node and Three-node Levels in the Long Delay Game	70
2.27	Means for all measures employed across the Training, One-Node and Three-Node Levels, across the Short Delay (n=11) and Long Delay (n=11) Games	70
2.28	Results from a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Total Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=22)	72

3.1	Number of attempts to pass each stage in the Relational Pre-Training and Testing procedure.	101
3.2	Number of attempts to pass each stage in the Relational Training and Testing procedure	102
3.3	Number of Correct Responses in the presence of each relational cue during the Game Phase	104
3.4	Mean Correct Responses across all four levels of the game (n=16)	107
3.5	Mean ratings of Positive Affect, Competence, Negative Affect and Impatience across all four levels in the gaming phase (n=16)	108
3.6	Results of a one-way Repeated Measures ANOVA examining the effects of Relational Complexity on Correct Responses, Positive Affect , Competence, Negative Affect and Impatience (n=16)	111
3.7	Number of attempts required by each participant to pass each stage in the Relational Pre-Training and Testing procedure	118
3.8	Number of attempts required by each participant to pass each stage in the Relational Training and Testing procedure	119
3.9	Total number of mouse clicks made in each level of the game by each participant	120
3.10	Mean Correct Responses across all across the No Delay, Half Delay and Full Delay levels (n=12)	122
3.11	Mean ratings of Positive Affect, Competence, Negative Affect, and Impatience across all four levels of the game (n=12)	128
3.12	Results of a one-way Repeated Measures ANOVA examining the effects of Network Latency on number of Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=12)	128
4.1	Mean Correct Responses, Positive Affect, Negative Affect and SCL, across all conditions in the experiment(n=10)	149
4.2	Slope and correlation co-efficient values for the best-fit line of mean SCL for each game (n=10)	160
4.3	Mean Correct Responses, Negative Affect and Positive Affect across the Slow Medium and Fast Games, with data from 2-stimuli and 6-stimuli games combined(n=10)	161
4.4	Mean Correct Responses, Negative Affect and Positive Affect across the 2-stimuli and 6-stimuli games, with data from the Slow, Medium and Fast games combined (n=10) .	163

4.5	Results of ANOVA examining the effects of Speed of Presentation and Stimulus Number on number of Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=10)	164
4.6	Mean Correct Responses, Positive Affect and Negative Affect across all six games played in Stage 1 (n=9)	184
4.7	Slope and correlation coefficients (r^2 . values) for games in Stage 1 that were later chosen to be played in Stage 2 and from those that were not chosen to be played in Stage 2 (n=9)	190
4.8	Slope and correlation coefficients (r^2 . values) for all games in Stage 1 that each participant rated highest for Positive Affect, and from all other games (n=9)	193
C.1	Sequence of games played by each participant in Experiment 4.2. Note that Stages 1 and 3 involved the presentation of all six games in randomised order, while the order of games presented in Stages 2 and 4 was determined by each participants' choice. . .	254

List of Figures

1.1	Representation of an Equivalence Relation	22
2.1	Matching-to-sample tasks for stimulus equivalence training (upper panel) and testing (lower panel).	34
2.2	Representation of one task in the training phase of the game.	37
3.1	Representation of the relational network trained and tested in Whelan and Barnes-Holmes (2004).	87
3.2	Representation of the relational network trained and tested in Experiment 3.1.	96
3.3	A screenshot of the user interface employed in the game phase of Experiment 3.1	98
3.4	Correct Responses across the SAME, OPPOSITE and Mixed SAME and OPPOSITE Levels for each participant	106
3.5	Figure 3.4 continued	107
3.6	Ratings of Positive Affect, Competence, Negative Affect and Impatience across the SAME, OPPOSITE and Mixed SAME and OPPOSITE Levels for each participant	109
3.7	Figure 3.6 continued	110
3.8	Correct Responses across the No Delay, Half Delay and Full Delay levels for all participants	123
3.9	Figure 3.8 continued	124
3.10	Ratings of Positive Affect, Competence, Negative Affect and Impatience across the No Delay, Half Delay and Full Delay Levels for all participants	126
3.11	Figure 3.10 continued	127
4.1	Mean Correct Responses across the six games(n=10)	150
4.2	Mean Positive Affect across the six experimental conditions(n=10)	150
4.3	Mean Negative Affect across the six experimental conditions(n=10)	151

4.4	Mean SCLs across the six experimental conditions(n=10)	152
4.5	Mean SCL for all participants for the duration of the 2-stimuli-slow game(n=10)	154
4.6	Mean SCL for all participants for the duration of the 2-stimuli-medium game(n=10)	155
4.7	Mean SCL for all participants for the duration of the 2-stimuli-fast game(n=10)	156
4.8	Mean SCL for all participants for the duration of the 6-stimuli-slow game(n=10)	157
4.9	Mean SCL for all participants for the duration of the 6-stimuli-medium game(n=10)	158
4.10	Mean SCL for all participants for the duration of the 6-stimuli-fast game(n=10)	158
4.11	Best-fit line for mean SCL recordings from each game (n=10)	160
4.12	Mean SCL data across all participants, across the Slow, Medium and Fast games. Data from the 2-stimuli and 6-stimuli games are combined to produce this graph (n=10)	162
4.13	Mean data across all participants across the 2-stimuli and 6-stimuli games. Data from the slow medium and fast games are combined to produce this graph (n=10)	163
4.14	Screenshot of the game choice screen presented to participants during Stage 2 and 4 of Experiment 4.2	172
4.15	Graphic representation of the number of participants who chose each game during Stages 2 and 4	174
4.16	Mean Correct Responses for each game across stages 1-3 (n=30)	175
4.17	Mean Positive Affect ratings for each game across stages 1-3 (n=30)	176
4.18	Mean Negative Affect ratings for each game across stages 1-3 (n=30)	177
4.19	The number of participants who chose to play each game in Stage 2 of Experiment 4.3	183
4.20	Mean Correct Responses across the six games played in Stage 1 (n=9)	184
4.21	Mean Positive Affect ratings across the six games played in Stage 1 (n=9)	185
4.22	Mean Negative Affect ratings across the six games played in Stage 1 (n=9)	186
4.23	Mean SCL recordings for all games in Stage 1 that were later chosen to be played by each participant in Stage 2 (n=9)	188
4.24	Mean SCL from every game in Stage 1 that was not chosen to be played in Stage 2 (n=9)	189
4.25	Best-fit line for mean SCL recordings for all games in Stage 1 that were later chosen to be played by each participant in Stage 2, and from those that were not chosen (n=9)	190
4.26	Mean SCL recordings for all games in Stage 1 that each participant rated highest for Positive Affect (n=9)	191
4.27	Mean SCL recordings for all games in Stage 1 other than those that each participant rated highest for Positive Affect (n=9)	192

4.28	Best-fit line for mean SCL recordings for all games in Stage 1 that each participant rated highest for Positive Affect, and from all other games (n=9)	193
4.29	Illustration of the predictive relationships between the dependent measures employed in Experiments 4.1-4.3	198
B.1	SCL Data from all games played by Participant 1 in Experiment 4.1	243
B.2	SCL Data from all games played by Participant 2 in Experiment 4.1	244
B.3	SCL Data from all games played by Participant 3 in Experiment 4.1	245
B.4	SCL Data from all games played by Participant 4 in Experiment 4.1	246
B.5	SCL Data from all games played by Participant 5 in Experiment 4.1	247
B.6	SCL Data from all games played by Participant 6 in Experiment 4.1	248
B.7	SCL Data from all games played by Participant 7 in Experiment 4.1	249
B.8	SCL Data from all games played by Participant 8 in Experiment 4.1	250
B.9	SCL Data from all games played by Participant 9 in Experiment 4.1	251
B.10	SCL Data from all games played by Participant 10 in Experiment 4.1	252

Chapter 1

Introduction

The computer games industry is currently the world's fastest growing entertainment industry, and has overtaken cinema box office sales in terms of revenue. "In the United States alone, the games industry reported about \$6.9 billion in sales in 2002, and sales increased to \$7 billion in 2003 and \$7.3 billion in 2004" (Davis, Steury, & Pagulayan, 2005, ¶ 1). In addition to these sales figures, many millions of people throughout the world pay monthly subscription fees, typically \$12 per month, for the privilege of playing games online with and against millions of other people. Revenues from such online gaming are estimated at \$1.5 billion in 2004 (Castronova, 2003) and some forecasters predict that the online gaming market will grow to \$13 billion by 2012 (Buckley, 2007).

Despite the enormity of this industry, and as result of its relative youth, little academic research yet exists on the activity of game playing. Aarseth (2001) suggests "we have a billion dollar industry with almost no basic research" (¶ 9) While scientific research into game play and game design has begun to take place, there is a lack of cohesion between the different strands of such research. Engineers investigate the engineering of games, social scientists examine the social impact of games on individuals and groups, Human-Computer Interaction (HCI) researchers look into the interaction of humans and computers while playing games. These approaches are disjointed and it is often difficult to translate results found through one type of investigation to another. The aim of this thesis is to bring a more systematic, experimentally sound and psychologically relevant research programme to bear on variables identified as important components of online game playing. In order to understand the task at hand, it is important to briefly consider some relevant issues arising from the existing literature on game playing.

1.1 What is a game?

In order to better understand computer games, it may be helpful to first look at the research on games and game playing in general. “It makes sense to look at computer games as being the latest development in a history of games that spans millennia” (Juul, 2003a, ¶ 39). There are numerous definitions of what constitutes a game, the majority of which appeal to enjoyment, (Malone & Lepper, 1987; Davis et al., 2005) repetition (Coyne, 2003) and competition (Morlock, Yando, & Nigolean, 1985; Vorderer, Hartmann, & Klimmt, 2003). Bernard Suits (1978) suggested that, “to play a game is to engage in activity directed towards bringing about a specific state of affairs, using only means permitted by rules, where the rules prohibit more efficient in favor of less efficient means, and where such rules are accepted just because they make possible such activity” (p.34). Salen and Zimmerman (2003) offer a similar, if simpler definition, “a game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome” (p. 96).

Juul (2003a) provides a particularly comprehensive and useful definition of a game. He proposes that there are six features which are both necessary and sufficient to define an activity as a game. These are as follows: 1) Games are rule-based, 2) Games have variable, quantifiable outcomes, 3) Different potential outcomes of games are assigned different values; Some outcomes are positive, while others are negative, 4) Players must exert effort in order to influence the outcome of the game, 5) The player must be attached to the outcome; An outcome designed as positive must result in happiness or satisfaction in the player who achieved this outcome, 6) Games must have negotiable consequences; The same game should be playable with the same rules, with or without real-life consequences. Juul proposes that only activities comprised of these six features can be described as games. Activities which are characterised by a number of, but not all of, these activities may resemble games, but should not strictly be considered games.

Gingold (2005), in his discussion of the computer game Wario Ware, points out “the essential elements of what makes a video game a video game” (¶ 13) are: Fiction, Goal, and Agency. Fiction refers to both the character the player is playing and their relation to the environment around them. For example, in *Doom III* (© id Software, 2004) the player is a US marine trapped on a scientific research station on Mars. Goal refers to the desired outcome of a play session. In *Doom III* the goal of a play session would be to stay alive and advance on to the next area of the research facility. Agency refers to the mechanism through which that goal can be met. In our example from *Doom III*, agency would refer to the action of aiming at shooting at the monsters while navigating through the various levels.

Unfortunately, both Juul and Gingold’s definitions preclude a number of very popular

open-ended simulation games such as *SimCity* (© Maxis, 2000), *The Sims* (© Maxis, 2000) and *Football Manager* (© Sports Interactive, 2006), as such games do not describe some potential outcomes as better than others. Rather, these games could be played continuously for years. There does not appear to be a traditional game which parallels the type of freedom and lack of conclusion found within open-ended computer-based simulation games, which poses the question of whether there is something fundamentally different between traditional card, board and dice games that have existed for hundreds of years and modern computer games.

1.1.1 Are Computer games different from traditional games?

While little research has been conducted over the years on the playing of traditional games such as chess, draughts, dice and board games, modern computer games researchers (Aarseth, 2001; Juul, 2003a; Woods, 2004) have recently examined the relationship between traditional games and computer games. Juul (2003a) has pointed out that games such as chess, solitaire and poker are ideally suited to playing on a computer. “It is. . . . one of the stranger ironies of human history, that the games played and developed over thousands of years have turned out to fit the modern digital computer so well” (¶ 49). Woods (2004) has pointed out that while computer games and analogue games appear quite different on first glance, many are structurally very similar. Juul (2003a) maintains that a computer is actually a better platform for games, both traditional and digital, as the computer can be programmed with the rules of the game so that players cannot make an illegal move. The computer can also keep track of the game state so that mistakes and cheating can be eliminated and players can then concentrate more fully on playing the game. Additionally, and perhaps more importantly, as players no longer have to concentrate on upholding the rules of the game, this allows a situation where players can begin playing a game without knowing the rules at all and perhaps play unaware of the potential outcomes of their actions.

1.1.2 What makes a game a good game?

Most game designers and theorists agree that “on the most basic level, the primary goal in a game is to be enjoyed” (Davis, Steury, & Pagulayan, 2005, ¶ 5). Put simply, good games are fun and intrinsically motivating (Malone & Lepper, 1987). However, many diverging accounts have been given as to what factors contribute to the ‘fun’ experienced by computer game players. Most accounts appeal to ‘challenge’ in one form or another. Other factors discussed include repetition, fantasy, narrative, flow, immersion, learning, usability and multiplayer interaction.

Usability may be described as one of the most basic elements in the creation of an enjoyable

computer game. Game designers focus much attention on the usability of the game so that players do not have to struggle with the interface, rather devoting their entire attention to overcoming the challenges laid forth in the game design. “Usability research can help identify problems or issues that block users from experiencing the fun of a game” (Davis et al., 2005, ¶ 2). A game with poor usability is very unlikely to invoke any enjoyment in those who choose to play it. However, usability in itself is not enough to constitute a fun game, as most word processors and other computer based work tools rate highly for usability, whereas a game that scores poorly for usability is very unlikely to be rated as enjoyable by players. Thus, a good usability rating may be viewed as a pre-requisite for a fun game.

Playability is another commonly discussed factor in contributing to the enjoyment of a game and is closely related to usability. However, the two concepts differ in a number of respects. Usability is a concept developed in HCI research for evaluating how easy to use a particular piece of software is. Concepts such as challenge and frustration are undesirable results in a usability analysis, while these are often vital elements of game play. “The goals of software productivity are to make the software interface easy to learn, use, and master, and somewhat oppose design goals for games, usually characterized as “easy to learn, difficult to master” (Desurvire, Caplan, & Toth, 2004, ¶ 1).

A closely related factor to playability is re-playability, which has been discussed at length by Ernest Adams (2001a). Adams states that “the single most important contributor to a game’s re-playability is its playability in the first place” (¶ 2). A game which is designed badly or not fun to play in the first place, will not be fun to play repeatedly. According to Adams, if a game is to be re-playable it requires a simple challenge and a well-designed user interface, so that a game task is reduced to the challenge and the means of completing it. “The most replayable games are also the smallest and cheapest to implement” (¶ 16). Games which are very re-playable are typically games in which a large amount of variation in gameplay is available. Classic arcade games such as pac-man offer the challenge of beating or finishing the game. However, once this has been achieved, there is little incentive to go back and play the game again, as there is very little variation available within the game play. Alternatively, multiplayer games will always be quite re-playable, as the variation in game play is created by opponents’ moves. Multi-player games such as Chess and Poker are hugely re-playable and have stood the test of time. Thus, re-playability may serve as a valuable factor in contributing to enjoyment in games.

1.1.3 Challenge and Problem-solving

The majority of those who have discussed the issue of what makes a computer game enjoyable to the player have appealed to ‘challenge’ in their explanations (i.e., Davis et al., 2005; Gingold, 2005; Kiili, 2005b; Malone, 1982; Morlock et al., 1985). Importantly, the idea of ‘challenge’ as appealed to by games researchers, appears to represent the problem solving involved in a game on either a basic stimulus-response or more complex verbal level. When researchers refer to the challenge presented to game players, this could be referred to as the problem presented that needs to be solved. For example, Malone (1982) declares that, “for an activity to be challenging, it needs to have a goal whose outcome is uncertain. . . . a challenging toy must either build in a goal or be such that users can easily create their own goals for its use” (p. 65). Gingold proposes that “in games you have to accomplish some goal, otherwise you lose” (§ 8). Both of these suggestions emphasise the goal state which must be achieved through some action of the player. Interestingly, a behavioural definition of problem solving contends that; “problem solving may be defined as any behaviour which, through the manipulation of variables, makes the appearance of a solution more probable” (Skinner, 1953, p.247). This behavioural definition of problem solving appears to parallel the suggestions of Malone (1982) and Gingold (2005) that games require players to take action in order to accomplish a goal state.

Game researchers have noted that the existence of a goal state is not enough in itself to define the challenge inherent in a game. (Juul, 2003b) proposed that in order for a game to be challenging, “no single option can be consistently superior to the others,” “the options should not be equally attractive,” and “the player must be able to make an informed choice” (§ 4). Again, this description appears to sit well with a Skinnerian definition of problem solving more generally; “when consequences are important and the probabilities of two or more responses are nearly equal, a problem must be solved” (Skinner, 1974, p.124). Juul’s definition of challenge in a game appears to require players to engage in problem-solving behaviour. Thus, in appealing to challenge as a key component in the enjoyment of a game, researchers such as Malone, Gingold and Juul, appear to suggest that the enjoyment of a game is derived from a player actively solving a problem that is posed during game play.

Some researchers have proposed that the process of learning contributes to the enjoyment experienced while playing a game. “When the game ceases to teach the player a new lesson, the game stops being fun” (Woods, 2004, § 32). Learning may be considered as part of the challenge of a game, as the player is constantly engaging in learning about the most effective ways to deal with the challenges posed in games. More specifically, both the overcoming of a sequence of challenges

and the learning of how to deal with a complex system may be viewed as specific types of problem solving. It seems that the activity of problem solving may be a major contributor to the enjoyment experienced by game players.

Vorderer, Hartmann and Klimmt (2003) note the importance of having challenges that are matched to the players own skill level. "Playing computer games is. . . . expected to be fun only if a sufficient portion of the competitive game situations is mastered by the player. . . . For this reason, many games allow for adjustments of difficulty levels in order to regulate the probability of success and failure in competitive situations according to the player's skill" (p. 3). Playing video games is expected to be enjoyable only if there are a sufficient number of successfully completed competitive situations. It appears that the inclusion of variable difficulty levels in many games may serve as an example of games designers allowing players to match the challenges of the game to their own skill level, and thus creating a broader appeal for their game.

Despite the attention given to the role of challenge in computer gaming, very few studies have empirically evaluated the relationship between challenge and enjoyment in computer game playing. Vorderer, Hartmann and Klimmt (2003) report one study that did examine the relationship between challenge and enjoyment. The experiment presented participants with verbal descriptions of game situations and asked them to rate these situations using a subjective scale. A linear correlation between challenge and enjoyment did not emerge from the results of this study. Indeed, players seemed to report greater enjoyment of a game in which they had a wide range of weapons to choose from than a game which involved more competitive, or challenging features. Thus, the authors' claims that a game must provide solvable problems are not supported. However, inadequacies with the experimental design may explain the unexpected outcome. Indeed, the authors note that, "future studies should attempt to replicate these findings by using real computer games instead of verbal game descriptions only" (p. 4). It is apparent that there is a lack of empirical studies that investigate the relationship between challenge and enjoyment.

Numerous other factors have been appealed to in explaining enjoyment in computer games. These include repetition (Coyne, 2003), fantasy and curiosity (Malone, 1982), immersion (Burke, 2005), flow (Jennings, 2002; Kiili, 2005a; Voiskounsky, Mitina, & Avetisova, 2004), consequences of play (Adams, 2001b), and principles of behaviour such as schedules of reinforcement (Loftus & Loftus, 1983). However, despite the recent increase in academic attention to computer games, "systematic quantitative methods for measuring and assessing the fun of a game are rarely employed in the games development process," (Davis et al., 2005, ¶ 7) or in computer games research for that matter. Discussions of enjoyment in games are generally anecdotal, qualitative or theoretical and

there is a lack of rigorous empirical work on the discipline in general.

1.2 Approaches to measuring enjoyment in computer games

Although there has recently been increased academic interest in computer games, the focus of work by many researchers is theoretical, qualitative or anecdotal. In addition, game designers typically work by rules of thumb and what research is conducted by games companies is not published. Thus, there exists a dearth of rigorous empirical research on computer gaming in general. Quantitative, experimental investigations into the factors contributing to an enjoyable computer game are uncommon. However, a small number of methods for measuring enjoyment in computer games do exist. Some of these have been developed by the computer gaming industry, such as usability testing (see Davis et al., 2005; Desurvire et al., 2004) playability testing (see Desurvire, Caplan, & Toth, 2004; Fabricatore, Nussbaum, & Rosas, 2002) beta testing and focus groups. Others have been developed by academic researchers, including behavioural measures, physiological recording and questionnaire based studies of measures such as *'flow'*.

1.2.1 Questionnaires

The majority of academic studies investigating enjoyment in computer games do so through the presentation of questionnaires after game play. However, there are relatively few studies which directly measure enjoyment in computer game playing through questionnaires. Often, a related attribute such as 'presence,' (i.e., participants subjective feeling of immersion in a virtual environment; see Krauss, Scheuchenpflug, Piechulla, & Zimmer, 2001; Witmer & Singer, 1998), or *'flow'* (an analysis of subjective positive experience based on pursuance of goals; see Voiskounsky et al., 2004) is investigated on the assumption that greater presence or greater flow correlates directly with greater enjoyment. However, this assumption itself is typically not tested.

One study that does examine enjoyment in computer game playing is reported by Prendinger, Mori and Ishizuka (2005). The researchers included a short five-item questionnaire in their study of frustration in computer game playing. Participants were asked to rate games that varied in terms of difficulty on five items that dealt with enjoyment, frustration and difficulty. In addition, questionnaire responses were correlated with Electrodermal Activity (EDA) and it was found that both ratings of frustration and EDA indicators of frustration were higher on the hypothesized most frustration inducing games. Thus, the questionnaire developed by Prendinger et al. (2005) may prove to be a valuable instrument in the investigation of both frustration and enjoyment in games.

In addition to those questionnaires developed specifically for the investigation of enjoyment in games, a number of instruments currently exist which have been validated by large samples and which could possibly be adapted for the purposes of measuring enjoyment in computer game playing. Stevens, Moget, DeGreef, Lemmink and Rispens (2000) have developed the Groningen Enjoyment Questionnaire (GEQ) which is a short ten-item questionnaire designed to measure enjoyment of leisure-time activities. Kahneman, Krueger, Schkade, Schwartz and Stone (2004) have developed the Day Reconstruction Method (DRM). This instrument is used to measure participants quality of life, by asking them to record events and rate them on a number of factors using a seven-point Likert scale. Importantly, the DRM provides a standardized method of evaluating specific recent experiences for their enjoyability, and thus may be adapted for the purposes of studying computer game playing experiences.

1.2.2 Physiological recording

Psychophysiological measurement is another potential outcome measure for use in quantifying enjoyment of computer games. Psychophysiology-based studies attempt to identify the underlying condition of the user in real-time during computer game play, concentrating on the measurement of either emotional states or cognitive activity. “The goal of physiological computing is to transform bioelectrical signals from the human nervous system into real-time computer input in order to enhance and enrich the interactive experience” (Allanson & Fairclough, 2004, p.857). There have been great advances in psychophysiological technology in recent decades which have allowed for greater clarity in the data obtained from such studies. A large number of methods, including electroencephalogram (EEG), electromyogram (EMG), electro-oculogram (EOG), electrocardiogram (EKG), respiratory measures, Electrodermal Activity (EDA), blood pressure and eye movement tracking have been investigated with respect to their suitability for providing computer input.

While psychophysiological methods have commonly been used for measuring general arousal levels and patterns of changes in arousal, until recently nobody had demonstrated which psychophysiological signals, which features of those signals, or which methods of classification could reliably differentiate emotions from each other (Vyzas, 1999). As one group of researchers put it; “how would we know whether we were measuring positive emotion such as fun, negative emotion such as frustration, or mental workload?” (Cahill, Ward, Marsden, & Johnson, 2001, ¶ 1). However, there has been some significant recent research in the field of ‘affective computing,’ or the psychophysiological recording of emotion (e.g., Allanson & Fairclough, 2004; Picard, 1997; Nasoz, Alvarez, Lisetti, & Finkelstein, 2003; Scheirer, Fernandez, Klein, & Picard, 2002; Vyzas, 1999).

Vyzas (1999) reports a study which demonstrated that it may be possible to distinguish between individual emotions using physiological data. EMG, BVP, EDA and respiration from chest expansion were recorded while an actress portrayed eight different emotions. All eight emotions were recognised from the physiological data at significantly higher than chance probabilities, demonstrating that there is significant information in physiological data for recognizing the emotional state of a person who is deliberately expressing a small set of emotions. Physiological recording methods have also been used to measure cognitive activity. Measures such as EEG and respiration rate are typically used to evaluate the level of difficulty a person is experiencing with a particular task. For example, Fairclough, Venables and Tattersall (2005) found that psychophysiological variables could be used quite reliably as a measure of mental effort. It appears that advances are being made in the identification of psychophysiological indicators of both emotional states and cognitive activity.

While much research has been conducted on the recording and analysing of emotion and cognition from physiological signals, little research has concentrated on the psychophysiological measuring of emotion and cognition during computer game play. However, Prendinger, Mori and Ishizuka (2005) report a study where EDA data was recorded while participants played a simple mathematical game. Delays, designed to frustrate the user, were inserted into the game and a virtual character interacted with the participants while playing the game. Prendinger et al. found that psychophysiological responses appeared to hold some promise for the investigation of emotional arousal in general. However, much further work must be conducted to investigate whether this method is suitable for the evaluation of enjoyment in computer games. Indeed, the EDA recording method will be of particular relevance to some of the aims of the current research (see Chapter 4).

1.2.3 Behavioural Measures

Many studies have been conducted in the past 15 years which have used computer games as a tool for the investigation of basic behavioural processes in humans. Although these studies have not set out to explicitly analyse or comment on the behaviour of computer game players *per se*, their work on such topics as schedules of reinforcement may nonetheless provide a valuable starting point for a behaviour analytic investigation of gaming. For example, DeHouwer, Beckers and Glautier (2002) used a computer game-like environment in which to evaluate human contingency learning. However, as the outcome of the game was pre-programmed and not controllable by the participant, this paradigm offers little insight into computer game playing.

Case, Ploog and Fantino (1990) investigated ‘observing behaviour’ in a computer game which was almost identical to one of the most popular games at the time the study was conducted. There

were a large number of different responses available to players and the reinforcers available were access to different parts of the game and a high score. Thus, the computer program employed was very game-like and the study was ideally suited to the investigation of computer game playing behaviour insofar as reinforcers of two kinds were under the control of the experimenters. However, the researchers focused on investigating observing responses, or responses that produce stimuli correlated with schedules of reinforcement but which have no effect on the occurrence of reinforcement. Little comment was made about game playing in general.

Raia, Shillingford, Miller and Baier (2000) used a computer game to investigate human sensitivities to variable ratio and variable interval reinforcement schedules. In another study, Leung (1989) employed a computer game in order to investigate preference for different schedules of reinforcement. Although these studies simply used computer games as a method of presenting basic behavioural experiments to human participants, findings of these experiments may hold some value for the analysis of simple games and simpler elements of more complex games.

Overall, the study of enjoyment in computer games lacks a cohesive theoretical background or methodology. A number of different approaches to the problem of measuring game play in all its dimensions have been developed, including psychophysiological recording of emotion and cognition, and self-report questionnaires measuring concepts such as flow and presence. However, few studies have provided empirical evidence as to what makes a computer game enjoyable. Most studies also use some element of the game playing in their evaluations. Factors such as high scores, reaction times and accuracy are used to back up results gathered through other methods. However, much less emphasis is placed on these behavioural measures than on other methods of recording.

1.2.4 Issues specific to Online Games

Online games are games where several persons interact simultaneously over networks such as the internet. Panterl and Wolf (2002) point out that competing with remote human players is typically considered more interesting and challenging than playing against computer opponents, as human opponents are typically more intelligent, spontaneous, and intuitive (see also Adams, 2001b; Mandryk & Inkpen, 2004; Manninen & Kujanpaa, 2005). Indeed, the appeal of online multiplayer games is growing exponentially, with some forecasters predicting that the online gaming market will grow to \$13 billion by 2012 (Buckley, 2007).

As the popularity of on-line multiplayer game-playing grows, network latency becomes an ever increasing problem (Pantel & Wolf, 2002). Network Latency can be defined as an expression of how much time it takes for a packet of data to get from one designated point in a computer network to

another. The assumption is that data should be transmitted instantly between one point and another; that is, with no delay at all. However, due to current technology, and the fact that the speed of light is the maximum information transfer speed possible (Sharkey, Ryan, & Roberts, 1998), it is not possible to transmit data instantly over large distances, as is required by on-line games.

In order to minimize the amount of data which must be communicated between users in an on-line multiplayer game, the virtual environment is usually replicated on the end-users computer and only the changes in game state, rather than the whole game state, are required to be communicated between users. However, due to network delays, these updates can take some time to reach each user and can in fact take different amounts of time to reach every user. Thus, the game playing experience is compromised in terms of responsiveness and consistency (Delaney, Ward, & McLoone, 2003). Responsiveness refers to how quickly the environment reacts to actions the user performs. Consistency refers to the need for each user in the application environment to see the same events at the same time; otherwise users will be reacting to events that have already changed the environment. The greater the distance between users of the game, the larger these delays will be and the greater the discrepancy between each users' representation of the game state (Vaghi, Greenhalgh, & Benford, 1999).

In terms of game playing, network delays mean that actions taken in the game environment can have unpredictable results. For example, in a first-person shooter style game, a player may see a direct shot at an opponent miss its target or may be hit by a shot which did not seem in danger of coming near them. Unfortunately the problem of network latency will be ever-present for the foreseeable future of on-line game playing, and as such it is important to investigate the impact such delays can have on the performance of multiplayer games and the attractiveness of these games for the human players (Pantel & Wolf, 2002).

There have been a small number of studies which have investigated the problem of network latency from the end-users perspective. Vaghi et al. (1999) developed a two-player virtual ball game where one of the players was subjected to an increasing amount of network delay. Participants were required to maneuver within the virtual environment and make contact with a ball in order to hit it over a central net and into an opponents' goal. They also had to maneuver within the environment in order to block balls hit by their opponents from going into their own goal. While one player received fast updates on game-state, the other player experienced delays in their interaction with both the environment and the other player. The extent of the delays experienced was increased gradually as the experiment progressed. Vaghi et al. found that delays of up to 150 milliseconds were unnoticeable to the players, while delays of up to 350 milliseconds meant that the game was erratic,

but still playable. Delays of up to 500 milliseconds proved very confusing for the players, while delays of up to 999 milliseconds meant that players struggled to have any sort of interaction with the ball.

Hashimoto and Ishibashi (2006) conducted a similar study using the popular rock-paper-scissors game. This study also found that as latency increased, subjective ratings decreased. Additionally, these authors found that most participants perceived unfairness when latency exceeded approximately 100 ms. Ishibashi, Nagasaka and Fujiyoshi (2006) report findings from a similar study, where participants perceived unfairness when the difference in network latency between two terminals exceeded approximately 100ms. Pantel and Wolf (2002) conducted a study involving a racing game and concluded that a delay up to 50 ms is uncritical for a car-racing game, while a delay of over 100ms may be acceptable for a first person shooter game. These studies provide a guide as to how much latency is acceptable within a game-playing environment.

Delaney, Meenaghan, Ward and McLoone (2004) also conducted an investigation into the effects of network latency on end users performance in a simple game playing environment. The authors intended to investigate the accuracy of end-user's performance on a game in the presence of network latency. However, they also assessed the effects of latency variation, also known as jitter, on users of distributed applications, thereby more accurately modelling the unpredictable delays experienced by on-line game players. The game involved an object which moved around a computer screen on one of a number of possible trajectories. As latency and jitter was increased, the representation of the object on the user's screen became less reliable with regard to the actual position of the object on its trajectory. Delaney et al. found that increases in both latency and jitter produced a decrease in user performance. Furthermore, they found that a combination of a medium amount of latency with a large amount of jitter was more disruptive than a large amount of latency. This suggests that the effects of jitter should be considered more closely, given that unpredictable variations in latency are more common in the experience of online gamers than a gradual increase in latency or a complete absence thereof.

The majority of studies on latency report engineering methodologies for reducing the impact of latency on game play. However, there have been a number of suggestions on how to minimize the interference of latency and jitter on game playing from an end-user, or psychological point of view. Vaghi et al. (1999) reported that players adapted to the latency with a number of coping strategies, which may be exploited by game developers. Players behaved more conservatively, maneuvered around the screen differently and played the game at a slower pace when latency was in place, whether they were aware of the delays or not. Vaghi et al. proposed that providing users with feedback as to the extent of delays currently being experienced by them may be a more useful

strategy of dealing with this problem than avoiding or covering up the latency.

Delaney et al. (2004) also suggest that feedback on the level of latency currently being experienced may prove helpful to game-players. In their study, a visual feedback screen was included which provided feedback on the accuracy of users' responses every time a mouse click was made. They report that as the latency and jitter was increased in their game, subjects relied more heavily upon this visual feedback system. Thus, it appears that latency and jitter impact significantly on participants' enjoyment of online games, and that it may be possible to combat latency through end-user, or psychological means.

1.3 The Beginnings of a Psychology of Gaming

Many diverse approaches have been taken to the study of computer game playing. These range from Human-Computer Interaction, Engineering, Economics and Anthropology to Media Studies. However, there is a lack of a sound psychological account of computer game playing. This is rather surprising, given the size and importance of the industry and the popularity of online gaming as a recreational activity. Psychological expertise in such areas as operant conditioning, complex behaviour and problem solving would seem ideally suited to the study of online game playing and yet have been rarely applied. A psychological investigation of computer game playing, unlike most other approaches to the topic, necessitates an end-user analysis. Such an investigation must begin with the behaviour of playing a game. This, in turn, must involve conceptualising the individual tasks involved in a game in psychological terms.

Some researchers, while not psychologists themselves, have recently advocated a behaviour-based psychological analysis of computer game playing. For instance, Jørgensen (2003) proposed that; "Player action is based upon a strategy that comes into being as a result of an interplay between the game layout, the players knowledge and beliefs, and the moves of a human or computer opponent" (p.5). However, it must be pointed out that the concepts of knowledge and beliefs, suggested by Jørgensen were used without technical, functional definitions.

Bauckhage, Thureau and Sagerer (2003) make a similar suggestion to that of Jørgensen, proposing that the actions of a player in a game are determined by a combination of their game state (health, ammunition, progression, etc.) and current environmental influence and can be predicted to some extent using this information. Indeed, Bauckhage et al. attempted to use behavioural principles to improve the Artificial Intelligence of simulated computer opponents in a game. Dixon, Malak and Khosla (2000) conducted a similar study. Although the results of these

exploratory studies were modest, and not as rigorous as a typical behaviour analysis experiment, nor as strictly adhering to the literature, they do suggest that a behavioural analysis of computer game playing may be useful for the design and development of computer games in future.

The current work will adopt a modern behavioural analytic approach to computer gaming, based on the concept of derived stimulus relations and many of the terms and concepts from Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001a). RFT is a behaviour analytic theory of human language and cognition, which evolved in the context of trying to deal with complex human behaviour. As such, it is ideally suited to analysing the complex behaviour involved in computer game play and may well extend this analysis beyond simple schedule based analysis to include the more complex aspects of game play mentioned earlier (e.g., verbal problem-solving). However, before discussing RFT in detail a brief overview of the behaviour analytic approach must be laid out.

1.3.1 The Science of Behaviour Analysis

Behaviour analysis is an environmentally based approach to psychology, which assumes that all psychological events are to be understood as interactions of organisms in and with historical and situational contexts (Hayes, 1993). Thus, the behaviour of any organism, at any time, is attributed to an interaction of the personal history of that particular organism with the particular context in which it is located. The approach can often seem counter-intuitive, as common sense explanations of behaviour such as states of mind, emotions and personalities are rejected in place of environmental accounts. Indeed, as all behaviour is explained in terms of interactions between an organism and its environment, the very idea of free will and autonomy itself is questioned. These issues are a consequence of the rigorous empirical approach adopted by behaviour analysts. Mental constructs such as states of mind, feelings, personalities and so on are rejected as explanatory tools for behaviour as they cannot be directly manipulated in an experimental setting. Only the environment in which behaviour occurs can be manipulated with any degree of rigour and accuracy, and thus only the environment is provided as an explanation for resulting behaviour.

Behaviour analysts are interested in the function of behaviour, rather than its structure or topography. When attempting to gain prediction and control over a particular behaviour, a behaviour analyst will conduct a functional analysis, which involves examining the antecedents and consequences of that behaviour. For example, consider a developmentally delayed child who occasionally bangs her head quite heavily against a wall. The behaviour analyst would observe the child in order to determine what environmental stimuli may have occasioned the episode, and also

what consequences the head banging behaviour may have brought about. Perhaps the child's caregiver interacted more affectionately with her after an episode of such self injurious behaviour. In such a situation, the head banging behaviour would appear to have the consequence of increased attention and affection from the caregiver. If this were the case, the behaviour analyst may advise the caregiver not to react to the problem behaviour, and would observe the effects this change in consequence had on the behaviour. If the behaviour did not extinguish, further recommendations would be made based on the observation of antecedents and consequences of the child's head banging behaviour. A behaviour analytic approach always focuses on the function of behaviour and seeks to identify manipulable antecedents and consequences in the environment, in order to achieve prediction and control over that behaviour.

Modern behaviour analysis is heavily indebted to the work of B.F. Skinner (i.e., Skinner, 1953, 1959, 1974) and his contemporaries, who pioneered the study of the role of context in behaviour. Skinner referred to his approach to the study of behaviour as 'Radical Behaviourism' in order to differentiate it from the work of early behaviourists such as Watson (see Catania, 1998). Radical Behaviourists exclusively studied the behaviour of animals such as rats and pigeons and devised many experimental designs and apparatus to do so. A typical study by Skinner or his contemporaries involved an animal being placed into a specially designed box containing a lever and a food dispenser. The experimenter set up a contingency where delivery of food was dependent on either a fixed amount of time or some behaviour produced by the animal. An experimenter kept a record of the behaviour of interest using a cumulative recording device (Skinner, 1959), while the animal would interact with its environment. Numerous such studies were conducted, the preparation of which varied depending on the behaviour of interest.

Radical behaviourists had some success in defining principles of behaviour from their work with animals. Work with schedules of reinforcement, and in particular, operant conditioning, has proven to be the most influential on current behaviour analysis. The concept of the operant is considered central to all behavioural accounts. Operant behaviour identifies a situation where, "consequences of behaviour may 'feed back' into the organism," and, "when they do so, they may change the probability that the behaviour which produced them will occur again" (Skinner, 1953, p.59). For example, a rat may engage in many behaviours while trapped in a cage. If one of these behaviours, such as pressing a lever, is followed by a favourable consequence such as the delivery of food, the probability of this behaviour occurring in future will have been altered. The term operant "emphasises the fact that behaviour operates upon the environment to generate consequences" (Skinner, 1953, p.65). Skinner specified an operant as a three term contingency consisting of an

antecedent, a response and a consequence. He suggested that this concept could be used to describe all complex behaviours of organisms. In an operant response, both an antecedent such as an environmental context or discriminative stimulus, and a consequence combine to produce behaviour in an organism. "In operant conditioning we 'strengthen' an operant in the sense of making a response more probable or, in actual fact, more frequent." (Skinner, 1953, p.65).

Empirical investigation of operant responding has led to technical definitions of terms such as positive and negative reinforcement, punishment and avoidance. For example, a reinforcer was defined as any stimulus, the presentation of which as a consequence of a response leads to an increase in responding in that particular context in future (Catania, 1998). Thus, building on the concept of the operant, many further advances were made by Skinner and his colleagues in providing a coherent account of the behaviour of organisms. Indeed, it appeared that the principles and methodologies developed by radical behaviourists could be applied to almost any complex human behaviour.

Importantly, the behaviour of computer game playing has been subjected to just such a basic behaviour analysis by Loftus and Loftus (1983). The authors propose a technical account of computer game playing, using concepts established in the behaviour-analytic literature. Concepts such as operant conditioning, extinction and schedules of reinforcement are used to analyse human computer game playing. Variable schedules of reinforcement are described as the explanation for addiction in computer game players and a comparison is drawn between a person playing the popular arcade pac-man and a rat in one of B.F. Skinner's classic behavioural experimental preparations. Loftus and Loftus' account is both rigorous and interesting and appeared valid at the time, when success at contemporary games was based primarily on reaction times and reflexes. Loftus and Loftus proposed that these games require little more than stimulus-response behaviour on the part of the user and thus, a basic behavioural analysis, focusing on schedules of reinforcement, may provide valuable insight into the playing of these games.

In addition to the analysis of simple games, low-level behavioural processes may provide an insight into players' continued playing of, or addiction to, more complex games. For example, the variable difficulty levels available in most modern games may be seen as a method of adapting the schedules of reinforcement inherent in a game in order to produce the most game-playing behaviour in the user. In effect, choosing a difficulty level may be classified as a response on a concurrent schedule, where the choice made is dependent on the player's experience with the game and their ability to gain points in the difficulty level chosen. However, as mentioned earlier, modern games involve more than simply repeated and fluid stereotyped responses to a limited number of stimuli,

they also involve problem solving. For example, popular strategy games such as *Command and Conquer* (© Westwood, 1995), *Total War* (© The Creative Assembly, 2000), and *Civilisation* (© MicroProse, 1991) require a player to not only fight battles with multiple units of different characteristics, on a number of different terrains, but also to build economies, military bases, towns, cities and empires. Many strategies can be adopted for pursuing such goals and the very structure of an army, when battle is required, is due entirely to choices made earlier in game play. Skinner devoted quite a considerable amount of interpretive research to understanding such problem solving humans, although he never conducted research using human participants. In his book *Science and Human Behavior* (1953), Skinner described how the complex behaviour of problem solving among others, could be subjected to a behavioural analysis. Particularly relevant for the current thesis is Skinner's definition of problem solving.

“A person has a problem when some condition will be reinforcing but he lacks a response that will produce it... solving a problem is, however, more than emitting the response which is the solution; it is a matter of taking steps to make that response more probable, usually by changing the environment.” (Skinner, 1974, p.123)

Skinner insists that the behaviour which solves a novel problem is not a brand new behaviour or insight, distinct from an organism's unique operant history, but is simply a novel arrangement of already established behaviours (i.e., “taking steps to make that response more probable;” Skinner, 1974, p.123). Skinner contends that such a process could be applied to many types of real world problem solving including those encountered in games.

Interestingly, Gingold (2005), in explaining the appeal of the game *Wario Ware* (Nintendo, 2003), echoes the foregoing process. Specifically, the game consists of a large number of simple mini-games that last approximately five seconds each, arranged in a sequential order. The player must quickly learn the rules of each mini-game to progress. At the end of each level, the skills learned in the preceding mini-games must be combined in order to pass a more complex game (i.e., chaining of previously learned simple behaviours). Gingold proposes that the process of gradually learning simple behaviours and combining these as the game progresses explains the appeal of the game. Thus, Skinner's approach to problem solving would appear relevant to at least some modern games.

Despite the success enjoyed by radical behaviourists in providing a coherent account of the basic behaviour of organisms, the analysis of language and complex human behaviours such as problem solving has not been as successful. Skinner's *Verbal Behavior* (1957) attempted to account for human language and cognition using the existing principles of behaviour analysis, applying that which had already been learned from studies with animals to the domain of human language and

cognition. Skinner (1957) defined verbal behaviour as any behaviour on the part of a speaker, reinforced through the mediation of a listener, who is trained by a verbal community so as to mediate such reinforcement. Language was explained through established principles such as operant conditioning, stimulus discriminations, schedules of reinforcement, and so on. However, as the majority of work in defining such concepts had been conducted with animals, this analysis could not rest entirely upon empirical evidence. New concepts, such as the mand, the tact, the echoic and the autoclitic were devised, based on those established in animal studies. Importantly, the success or failure of the account would rest upon the empirical investigation of these concepts.

Unfortunately, in the years since Skinner's account was developed very little research has been conducted to examine just these types of complex behaviour. As a result, some have critiqued his account on the grounds that it is difficult to apply in practice, due to his definition of verbal behaviour not being a functional one (but see Cooper, Heron, & Heward 2006 for a discussion of how Skinner's analysis may still prove valuable and practical). Behaviour analysis holds as a fundamental tenet that all behaviour is an interaction of the personal history of a particular organism with the particular context in which it is located. However, Skinner defined verbal behaviour as any behaviour on the part of a speaker reinforced through the mediation of a listener who is trained by a verbal community so as to mediate such reinforcement. Thus, Skinner's analysis is not based on the specific aspects of an individual organisms history, but on that of another organism; the listener (Hayes & Barnes-Holmes, 2004). It would appear that an organism is behaving verbally only if a listener is listening verbally. Modern behavioural researchers (Hayes et al., 2001a) have argued that an analysis of the verbal behaviour of an organism must rest solely on the history and context of that organism while behaving.

In addition, Skinners definition of verbal behaviour has been criticized for being too broad. Hayes and Barnes-Holmes (2004) have proposed that Skinners definition of verbal behaviour is, "so broad as to include virtually all animal operant behavior in traditional behavior analytic research," (p. 218), due to the role of the listener in Skinner's account. For example, consider an experimental preparation where a pigeon must peck a key five times to receive a food pellet. This example fulfills Skinner's definition of verbal behaviour, as the key pecking response of the pigeon (the "speaker") is reinforced through the mediation of an experimenter (the "listener"), who has been trained to do so by the verbal community of behaviour analysts. Thus, Skinners definition does not distinguish verbal behaviour from other forms of social behaviour (Barnes-Holmes, Barnes-Homes, & Cullinan, 2000; Chase & Danforth, 1991).

Aside from any conceptual problems with Skinner's account of human verbal behaviour, the

success of his approach would ultimately rest with the pragmatic truth criterion of ‘successful working;’ whether or not it would lead to a generative research programme and greater prediction and influence over future behaviour. While some studies have successfully adopted Skinner’s (1957) approach to verbal behaviour in accounting for specifics of human language and cognition (e.g., Boe & Winokur, 1978; Lamarre & Holland, 1985; Lee, 1981; Lee & Pegler, 1982; Salzinger, 1958) “most behavioral researchers seem to agree that a relative dearth of empirical work was generated by Skinner’s approach” (Hayes, Blackledge, & Barnes-Holmes, 2001b, p. 10).

Empirical problems have also been created by the broadness of Skinner’s definition of verbal behaviour. Hayes et al. (2001b) argued that, because Skinner’s definition does not distinguish between verbal behaviour and other forms of social behaviour, “any attempt to apply the analytic categories described in the book (*Verbal Behavior*) lead basic behavior analysts inexorably back to what they were already doing in the [animal] laboratory” (p. 15). Furthermore, they argued that; “the book did not lead to a progressive research program that raised a large set of new and important questions about language” (Hayes et al., 2001b, p.11). In effect, Skinner’s account of language and complex behaviours such as problem solving was not sufficient to deal with complex human behaviour such as generative grammar, novelty and complex problem solving.

While the literature does not suggest any inherent flaws in the Skinnerian approach to complex behaviour, it simply appears that it is not progressive enough a program to generate fruitful experiments on complex behaviour. As such, it may prove more useful to pursue a more generative research programme in approaching complex human behaviours such as language, problem solving and computer game playing. However, despite the criticisms levelled against *Verbal Behavior* by modern behaviour analysts, it remains a highly influential text, which may still provide some valuable insight into human language and complex behaviour. Indeed, Barnes-Holmes, Barnes-Holmes and Cullinan (2000) suggested that combining Skinner’s work with a modern behavioural account of human language and cognition called Relational Frame Theory (RFT) would help to develop a clear and useful research agenda into complex human behaviour. In fact, Barnes-Holmes et al. (2000) aimed to develop a modern, coherent, naturalistic and purely functional-analytic understanding of human language that would provide a powerful challenge to the many non-behavioural approaches within psychological literature today by combining these two approaches. Murphy, Barnes-Holmes and Barnes-Holmes (2005) reported success in the use of this RFT/Skinnerian hybrid in an attempt to improve language deficits in a sample of autistic children. Thus, far from representing a break from Skinner’s *Verbal Behavior*, RFT builds upon the Skinnerian analysis to provide a more comprehensive approach to complex behaviour.

1.3.2 A Derived Relations Approach to Complex Behaviour

The core new concept of the post-Skinnerian approach that has made a novel analysis of human language and complex behaviour possible is that of Stimulus Equivalence. Stimulus Equivalence emerged from work conducted by Murray Sidman (Sidman, 1971; Sidman & Tailby, 1982; Sidman, 1994) on teaching developmentally delayed children to read using a matching-to-sample task procedure. A matching-to-sample task is a situation where a participant is presented with a sample stimulus and two or more comparison stimuli. The participant must choose which of the comparison stimuli ‘goes with’ the sample stimulus. This methodology was initially used by Sidman (1971) to teach reading comprehension to a severely developmentally delayed youth. The experimental preparation involved spoken words, pictures and written words as stimuli. Auditory stimuli were presented by the experimenter and the subject was required to choose the appropriate written word or picture that “goes with” the auditory stimulus. Appropriate responses were reinforced with both candy and money. Sidman had remarkable success in teaching written word comprehension to his subject. The participant was able to successfully identify the correct pictures when presented with a large number of written words, something he had been unable to do before Sidman’s intervention.

Sidman recognized that the matching-to-sample task was a useful and efficient methodology for teaching reading comprehension, as his participant not only responded correctly to relations that were taught explicitly during the matching-to-sample tasks, but also responded correctly on tasks which had not been directly trained. For example, if the participant was trained to choose the written word ‘horse’ when presented with the spoken word ‘horse,’ and then choose a picture of a horse when presented with the spoken word ‘horse’, the children were then able to successfully say the spoken word ‘horse’ when presented with the written word ‘horse;’ say the spoken word ‘horse’ when presented with the picture of a horse; choose the written word ‘horse’ when presented with the picture of a horse; and were able to choose the picture of a horse when presented with the written word horse; all without direct training. During the 1971 experiment, it took 15 hours of direct training for the participant to reliably match 20 spoken words to their written counterparts. However, once the participant had reached this level of performance, he was able to match the 20 written words to the relevant pictures, without having been directly taught to do so. Sidman (1971) proposed that his participant’s performances were as a result of the stimuli in his experiment becoming equivalent to each other through the matching-to-sample task. For example, the spoken word horse became equivalent to the written word horse and thus the subject responded to the written word horse, as if it was the spoken word horse. However, it was not until Sidman’s two 1982 papers (Sidman, Rauzin, Lazar, Cunningham, Tailby, & Carrigan, 1982; Sidman & Tailby, 1982)

that the full implications of this suggestion were discussed.

The research conducted by Sidman and his colleagues between 1971 and 1982 led them to conceptualise stimulus equivalence not only as a basis for reading comprehension, but as, “a behavioural basis for everyday correspondences between words and things, between what we say and what we do, and between rules and contingencies” (Sidman, 1994, p.123). Stimuli which were related to each other through equivalence were said to participate in stimulus classes and members of such classes were described as functionally substitutable for each other. In order to investigate the phenomenon further, a number of technical terms, based on the mathematical concept of equivalence, were devised to describe the pattern of behaviour exhibited by participants in Sidman’s equivalence experiments. These were deemed necessary requirements for the demonstration of an equivalence relation, and were labeled reflexivity, symmetry and transitivity.

Reflexivity relations require that stimuli related to each other must show the same relation to themselves. For instance, in a learning exercise a child may be taught to match a picture of a horse to the word horse. However, the child must also be able to match the word with itself and the picture with an identical one upon presentation.

Symmetry requires that relations that are taught must be reversible without explicit training. If a relation is symmetric, then that which has been taught as a sample should function as a comparison, and that which has been taught as a comparison should function as a sample (Barnes, Browne, Smeets, & Roche, 1995). Having learned to choose the word horse as a comparison when presented with the picture of a horse as a sample, the participant must choose the picture of the horse as a comparison when they are presented with the word horse as a sample. Transitivity requires that when a stimulus (A) is related to a second stimulus, (B) and A is also related to a third stimulus (C), then B is related to C (Barnes, 1994). For instance, if a participant is trained to relate the English word ‘one’ with the digit ‘1’ and also with the German word ‘ein’, then they must relate the digit ‘1’ to the German word ‘ein’ in order to pass the transitivity test. Therefore, a number of untaught, derived performances are demonstrated in the formation of an equivalence relation, both through symmetry and transitivity.

See Figure 1.1 for an illustration of an equivalence relation. The solid lines represent trained relations between stimuli A1 and B1, and also A1 and C1. The dotted lines between B1 and A1, and between C1 and A1 represent the derived symmetrical relations. The dotted lines between B1 and C1 represent the derived equivalence relations between these two stimuli.

Importantly, Stimulus Equivalence (SE) provided a conceptual framework and methodology for studying human language and complex behaviour within the behaviour analytic tradition. The

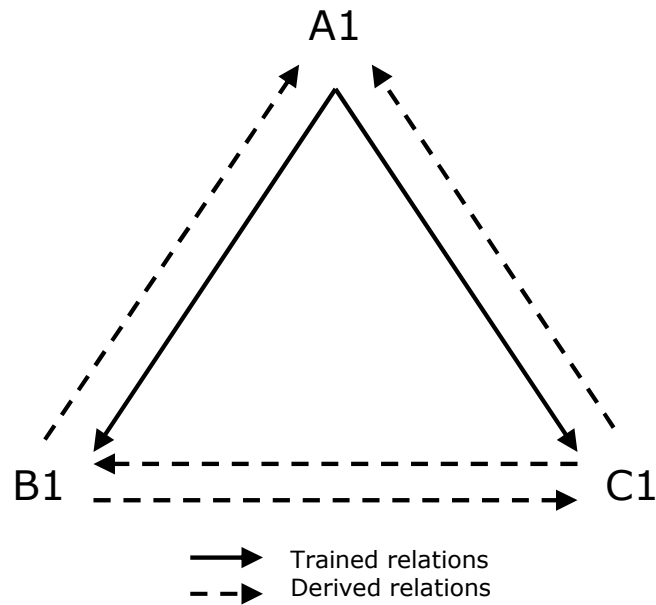


Figure 1.1: Representation of an Equivalence Relation

concept of SE has led to a generative and fruitful research programme into human verbal behaviour and associated processes (e.g., problem solving). “Derived stimulus relations, such as stimulus equivalence, have proved very exciting to behavioral researchers because their emergence is not predicted by traditional behavioral accounts, and they appear to parallel many natural language phenomena” (Dixon, Dymond, Rehfeldt, Roche, & Zlomke, 2003b, p.131). Indeed, Barnes (1994) and Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan and Leader (2004) present a large amount of diverse behaviour analytic research that provide evidence to support the claim for a close interrelation between stimulus equivalence and human language.

It was necessary to explain the philosophical and historical foundation of behaviour analysis in order to clarify the reasons for the adoption of the current research strategy. It is important to understand that this approach does not attempt to identify extant processes at work in the heads of individuals, but attempts to understand behaviour by gaining experimental control over it and identifying analogous contingencies at work in the natural environment. Secondly, the terminology of behaviour analysis is critical to the procedures that will be employed in subsequent chapters. Thirdly, the review of behaviour analysis as a discipline places any contributions made by the current research into the relevant context.

1.3.3 Relational Frame Theory: Extending the Derived Relations approach to Complex Behaviour

Relational Frame Theory (RFT) is a modern behavioural theory of language and cognition, which suggests that stimulus equivalence is not the only form of derived relational responding demonstrated by humans, but one of many (Hayes, Barnes-Holmes, & Roche, 2001a). A number of empirical studies (e.g., O'Hora, Roche, Barnes-Holmes, & Smeets, 2002; Roche, Linehan, Ward, Dymond, & Rehfeldt, 2004; Steele & Hayes, 1991; Whelan & Barnes-Holmes, 2004; Whelan, Barnes-Holmes, & Dymond, 2006) have demonstrated that participants can respond consistently in accordance with relations such as Similarity, Opposition, Difference More Than/Less Than, and other such comparisons. Importantly, these performances cannot readily be viewed simply as equivalence responding (Dymond & Barnes, 1995). RFT has developed new terminology to describe the processes involved in derived relational responding. These are Mutual Entailment, Combinatorial Entailment and Transformation of Function.

Mutual entailment: If a stimulus A bears a relationship to another stimulus B, then a further derived relation between B and A is mutually entailed. The type of relation mutually entailed depends upon the nature of the relation between A and B (Hayes, 1994). For instance, if the stimulus A bears an equivalence or SAME relation to the stimulus B, then the relation "B is the same as A" is mutually entailed. Thus, relations mutually entailed from equivalence relations are always equivalence relations. However, this is not always the case when alternative relations are trained. For example, if the stimulus A bears a relation of comparison to the stimulus B (e.g., A is more than B), then the relation "B is less than A" is mutually entailed. In other words, the derived response may be functionally distinct from the trained response. The distinction is even more apparent when we consider combinatorial entailment.

Combinatorial entailment: If a stimulus A bears a relation to another stimulus B, and B bears a further relation to a stimulus C, then a relation between A and C is combinatorially entailed. The nature of the combinatorially entailed relation depends entirely on the nature of the trained relations. As with mutual entailment, when the relation trained is an equivalence relation, the relation entailed is always an equivalence relation. For example, if A is the same as B and B is the same as C, then A is the same as C. However, when relations other than those of equivalence are trained, quite complex relations may be entailed. For example, if A is the OPPOSITE of B and B is the OPPOSITE of C, then a SAME relation between A and C is derived by combinatorial entailment (i.e., A is the SAME as C). Importantly, RFT suggests that relational responding in accordance with a frame in which the derived response is functionally distinct from the trained response constitutes a

more complex relational activity than responding in accordance with a frame in which the derived response is functionally identical to the trained response. Thus, responding to the opposite of an opposite would represent a more complex relational activity than an equivalence response.

Transformation of function: If a stimulus A is related to another stimulus B, and A acquires a psychological function, then in the appropriate context the stimulus functions of B will be transformed in accordance with the A-B relation. For example, if A is "more than" B, and A elicits fear, then B will produce less fear than A. RFT suggests that all derived relational responding involves a transformation of functions to some extent, as even the ability of a subject to point to a comparison that previously functioned as a sample during equivalence training, involves a transformation of the functions of that stimulus (Dymond, Roche, & Rhefeldt, 2005).

The idea that derived relational responding (DRR) itself constitutes generalized operant activity is pivotal to Relational Frame Theory (see Hayes et al., 2001a; Roche et al., 2004). It is suggested that derived relational responding constitutes an operant response class, in that it is created by a history of reinforcement across multiple exemplars, and once established, any stimulus event, regardless of form, may participate in the relational frame of equivalence (Barnes-Holmes & Barnes-Holmes, 2000). Limited research has been reported that demonstrates the sensitivity of derived relational responding to reinforcement contingencies (e.g., Healy, Barnes, & Smeets, 1998; Healy, Barnes-Holmes, & Smeets, 2000). In addition, some longitudinal research has tracked the emergence of derived relational responding in young children (e.g., Barnes-Holmes, Barnes-Holmes, Roche, & Smeets, 2001d,e; Lipkens, Hayes, & Hayes, 1993). However, the precise details of the history required to produce derived relational responding are not crucial to the RFT position on the generalized operant. Rather they are posed as important empirical questions (see Hayes & Wilson, 1996).

RFT extends the Stimulus Equivalence account of derived relational responding in that it both accounts for the performances of participants who consistently respond in accordance with relations other than that of equivalence, and through defining derived relational responding as a generalised operant, explains how such complex behaviour can develop.

1.3.4 How are derived relations relevant to computer gaming?

The RFT approach seems to allow for a description of many computer games in a technical manner and in a way that would appear to have face validity. For instance, consider a game that requires a player to react to two rapidly approaching characters (let us call them A1 and A2) on a computer screen and in which points are earned by responding to one character with a space-bar

press and the to the other with a mouse-click. One feature of this game is that the two characters involved are necessarily related to each other, in this case by a relation of opposition or difference (i.e., they require a different response). Indeed, it is critical that the characters are seen as different and not equivalent if the player is to achieve a high score and advance to the next level of the game.

Another relational feature of this hypothetical game is that the characters become cues (discriminative stimuli) for earning points by making the various appropriate operant responses required in the game. Behavioural research shows that humans will spontaneously form verbal relations between the stimuli under these conditions and are capable of privately or publicly verbalising these relationships (see Lovibond, 2004). It is this ability to verbalise relations between various actions, and between action and its consequences that behaviour analysts refer to as “knowledge” (Barnes-Holmes et al., 2001b). A detailed account of the RFT analysis of rule generation and following is beyond the scope of the present research (but see Hayes, 1989). Suffice to say at this point, however, that from an RFT perspective, the ability to state environmental contingencies (e.g., the conditions necessary to score points in a game) provides the basis for self-rule generation and effective rule-following. Thus, RFT provides a technical nomenclature with which to understand how individuals make choices in their responses to game characters while engaging in on-line gaming.

Of course games are typically more complex than described in the foregoing example. Imagine, for instance, a game in which additional characters are introduced. These new ones look like the original (A) stimuli, but differ along some physical dimension (i.e. they are bigger, longer, or different in hue). Participant’s responses to such characters may be accounted for by the principle of stimulus generalisation. Stimulus generalisation has been defined as, “the spread of the effects of reinforcement (or other operations such as extinction or punishment) from one stimulus to other stimuli differing from the original along one or more dimensions” (Catania, 1998, p.391). As the new characters in the game share physical similarities to the original characters, we would expect the appropriate responses to transfer to these new characters with little difficulty. However, there are not many modern games that rely on stimulus generalisations as part of game-play. Such a design is more typical of pioneering work in the early days of software engineering. Indeed, simpler psychological processes such as stimulus generalisation and simple conditioned reflexes may provide us with some understanding of the appeal of more rudimentary games such as *Pong*, *Pac-man* and *Space Invaders* (Loftus & Loftus, 1983). Nevertheless, stimulus generalisation is one simple process that may be involved in some aspects of computer games and is easily accounted for using a traditional behavioural principle.

Now imagine that even more characters are introduced to our simple game. These characters must be responded to, not based on their appearance, but on their relationship to the original characters. Let us call the two original characters A1 and A2. Two new characters, labelled B1 and B2, are introduced as allies of A1 and A2, respectively. Now two further characters, C1 and C2, are introduced as allies of B1 and B2, respectively. These relationships can be established most easily in the game through instructions, as they often are in complex strategy games, or simply through interaction with the game (i.e., trial and error). The important point is that no relation between the A and C characters is explicitly specified at any stage. Nevertheless, most game players will have little difficulty responding to the C characters appropriately. This, in effect, is derived relational responding, or more specifically, the transfer of the response functions of the A stimuli to the C stimuli. Of course real-world games involve the foregoing features wrapped in impressive presentation formats and a range of clever graphics, but at a technical and psychological level we can conceive of at least some aspect of gaming as involving derived relational responding. Indeed, the purpose of this research is to test this very approach and its effectiveness as a paradigm for the psychological analysis of game playing.

Chapter 2

Experiments 2.1-2.4

2.1 Experiment 2.1

Multi-player online gaming is an increasingly popular pastime in the 21st century (Davis, Steury, & Pagulayan, 2005). Despite this popularity, however, as yet relatively little scientific psychological research has been conducted into this phenomenon. The current study represents an initial attempt to analyze online gaming from a modern behavioural perspective by modeling two apparently central features of the gaming experience; *network delay* and *complexity* and quantifying the main interaction effects of these two variables on both gaming performance and self-reported satisfaction with the gaming experience.

The first of the two features of online gaming examined in this chapter is network delay, which is defined technically in terms of the time taken for a packet of data to travel from one designated point in a network to another (e.g., over the internet). The ideal scenario for game playing is that data should be transmitted instantaneously between two or more online gaming stations. However, due to technological limitations, psychologically significant amounts of time are often required for data transmission, producing noticeable delays between the response of the player and the reaction of the game software, thereby compromising the game-playing experience. The issue of network delay can perhaps be described as the greatest challenge to both the engineering of on-line games as well as the development of a coherent psychological account of on-line game playing processes and experience (see Delaney, Meenaghan, Ward, & McLoone, 2004; Vaghi, Greenhalgh, & Benford, 1999).

The second potentially psychologically important feature of online game playing and of computer gaming more generally is game complexity. The link between game complexity and the

enjoyment experienced by a player has been emphasized by many game designers and theorists (i.e., Gingold, 2005; Malone, 1982; Morlock et al., 1985; Vorderer et al., 2003). Malone (1982) proposed that; “a good game should be easy to learn, but difficult to master.” In a similar vein Vorderer, Hartmann and Klimmt (2003) pointed to the importance of having challenges that are matched to the player’s own skill level. In their words;

“Playing computer games is expected to be fun only if a sufficient portion of the competitive game situations is mastered by the player. For this reason, many games allow for adjustments of difficulty levels in order to regulate the probability of success and failure in competitive situations according to the player’s skill.” (p.3)

Thus, the inclusion of variable difficulty levels in many popular games allows players to match the complexity of the game to their own skill level, thereby creating a broader appeal for the game. However, while a great deal of importance is placed on complexity as a feature of game playing by games designers and theorists, a technical psychological account of the effects of complexity on game playing is absent from the literature.

In approaching the problem of analysing on-line game playing from a behavioural perspective there is little to no research base to serve as a guide for experimental methods or on definitional issues. Specifically, even the loosest psychological definition of a game appears to be unavailable in the literature on human-computer interaction, ergonomics, problem-solving and even complex cognitive processes more generally. Some researchers have attempted to study game playing at a qualitative (Klein, Moon, & Picard, 2002; Williams & Clippinger, 2002) and a psycho-physiological (Mandryk & Inkpen, 2004; Prendinger, Mori, & Ishizuka, 2005) level. Others have employed theoretical concepts such as *flow* (Voiskounsky, Mitina, & Avetisova, 2004) and *presence* to understand effective game playing. However, few of these studies have provided a scientifically-based account of the development of effective game playing skills, let alone a technical definition of what will constitute a game for research purposes.

Interestingly, one recent behavioural account of complex human functioning, known as Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche 2001a) appears to apply to precisely those skills involved in complex on-line games, and may serve as a guide to understanding the psychology of game playing at a technical level. Relational Frame Theory is a behaviour analytic account of human language and cognition, which evolved in the context of trying to explain complex human behaviour. As with all behavioural approaches, RFT assumes that all psychological events are to be understood as interactions of organisms in and with historical and situational contexts (Hayes, 1993).

The central psychological phenomenon of interest to relational frame theorists, and the one most obviously implicated in the explanation and analysis of complex behaviour, is the phenomenon known as *derived relational responding*, the simplest example of which is *stimulus equivalence*. The concept of stimulus equivalence evolved from research conducted by Sidman (1971; see also Sidman & Tailby, 1982; Sidman, Rauzin, Lazar, Cunningham, Tailby, & Carrigan, 1982), who reported that training a series of related conditional discriminations in developmentally delayed humans could result in the demonstration of a number of untrained or derived conditional discriminations systematically related to those trained. The procedure that Sidman used to demonstrate stimulus equivalence was the matching to sample (MTS) procedure. In a typical MTS task a participant is presented with a sample stimulus and two or more comparison stimuli and must choose the correct comparison stimulus for that sample stimulus. For example, Sidman showed that once participants had been explicitly taught to choose an arbitrary stimulus “B” in the presence of an arbitrary stimulus “A” and also to choose a third arbitrary stimulus “C” in the presence of stimulus “B,” then a number of untrained responses emerged as follows: Choosing “A” in the presence of “B,” and “B” in the presence of “C,” referred to by Sidman as symmetry; choosing “C” in the presence of “A” (referred to as transitivity) and “A” in the presence of “C,” (combined symmetry and transitivity). Sidman referred to the latter pattern involving both symmetry and transitivity as stimulus equivalence, because according to his analysis, if a participant demonstrated such a pattern it suggested that the participant was treating the stimuli as being functionally identical or equivalent to each other.

The concept of stimulus equivalence has generated much empirical work and conceptual discussion among behaviour analysts. In particular, the related effect known as a transfer of functions appears to have significant explanatory power in accounting for the acquisition of complex behaviour patterns typical in phobias and other apparently unlearned behaviours. For instance, imagine a human participant that is exposed to the equivalence training procedure described above. Now imagine that the A stimulus is subsequently paired with brief presentations of electric shock. When this participant is then presented with the C stimulus, they produce a spontaneous physiologically measurable fear response without instruction. In effect, the fear eliciting function of the B stimulus transfers to other members of the derived equivalence relation (see Dougher, Auguston, Markham, Greenway, & Wulfert, 1994). This effect has been used to account for a wide variety of responses previously outside the remit of behavioural psychology (see Wilson, Hayes, Gregg, & Zettle, 2001). Thus, derived relational responding and related effects appear to represent not only a basis for reading comprehension, but, “a behavioural basis for everyday correspondences

between words and things, between what we say and what we do, and between rules and contingencies” (Sidman, 1994, p.123). In other words, stimulus equivalence provides a conceptual framework and methodology for the empirical study of complex behaviour in humans.

Relational Frame Theory has expanded on the work done by Sidman, in that a larger number of relations (called relational frames) are appealed to in order to explain human language and cognitive abilities. Recent empirical research has demonstrated responding in accordance with a large number of relations, such as Opposite, Difference, (Steele & Hayes, 1991) and More Than and Less Than (O’Hora, Roche, Barnes-Holmes, & Smeets, 2002). Moreover, RFT has taken an explicit interest in the study of relations among relations, networks of which characterize what we might call metaphor and simile (Stewart, Barnes-Holmes, S.C, & Lipkens, 2001; Stewart, Barnes-Holmes, Roche, & Smeets, 2002), meaning (Barnes-Holmes, D. O’Hora, Roche, Hayes, Bisset, & Lyddy, 2001a; O’Hora, Barnes-Holmes, Roche, & Smeets, 2004) mathematics (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2001c), and rule following (Barnes-Holmes, Hayes, & Dymond, 2001b). This final application of the theory is what makes it particularly suited to studying complex behaviours, such as game playing, that appear to involve rule following and relating stimuli to each other while assessing the relations between objects and events in a problem solving context (Hayes, Gifford, Townsend, & Barnes-Holmes, 2001c).

Many on-line games, in particular those that are known as strategy and role paying games, appear to involve just those sorts of complex behaviours that are of interest to RFT (e.g., rule following, relating stimuli to each other and problem solving more generally). Several popular modern games require players to devise strategies through solving problems posed in the game, and to devise self-directed rules in order to employ those strategies in order to earn high scores and progress to further levels in the game. For example, consider the hugely popular genre of strategy games including titles such as *Command and Conquer* (© Westwood, 1995), *Total War* (© The Creative Assembly, 2000), and *Civilisation* (© MicroProse, 1991). These games require a player to not only fight battles, but to build economies, military bases, towns, cities and empires. Many strategies can be adopted for pursuing such goals and the very structure of an army, when battle is required, is due entirely to choices made earlier in game play. Management of resources and anticipation of enemy behaviour is as important in such games as fighting battles. Thus, understanding the developing chains of responses that are required to play these games well, and the attendant cognitive processes will be no easy task. Indeed, the current research venture into this new and challenging research field must begin with a virtual *tabula rasa*. Nevertheless, RFT would appear to represent a modern technically sophisticated and conceptually coherent framework within

which to begin such an endeavour. While other researchers may begin elsewhere in the study of complex gaming behaviour, it seems intuitive that the relational features of games are the most amenable to immediate study. Moreover, a systematic manipulation of these features would appear to be the best starting point in any attempt to gain experimental control over game complexity.

Using RFT we can conceive of at least some aspect of gaming as involving derived relational responding and these features are subject to experimental manipulation. For instance, consider a game that requires a player to react to two rapidly approaching characters (let us call them A1 and A2) on a computer screen and in which points are earned by responding to one character with a space-bar press and the to the other with a mouse-click. One feature of this game is that the two characters involved are necessarily related to each other, in this case by a relation of opposition or difference (i.e., they require a different response). Indeed, it is critical the characters are responded to as non-equivalent rather than equivalent if the player is to achieve a high score and advance to the next level of the game.

Another relational feature of this hypothetical game is that the characters become cues (discriminative stimuli) for earning points by making the various appropriate operant responses required in the game. Behavioural research shows that humans will spontaneously form verbal relations between the stimuli under these conditions and are capable of privately or publicly verbalising these relationships (see Lovibond, 2004). It is this ability to verbalise relations between various actions, and between action and its consequences that behaviour analysts refer to as “knowledge” (Barnes-Holmes, Hayes, & Dymond, 2001b). A detailed account of the RFT analysis of rule generation and following is beyond the scope of the present paper. Suffice to say at this point, however, that from an RFT perspective, the ability to state environmental contingencies (e.g., the conditions necessary to score points in a game) provides the basis for self-rule generation and effective rule following. Thus, RFT provides a technical nomenclature with which to understand how individuals make choices in their responses to game characters while engaging in on-line gaming.

Of course games are typically more complex than described in the foregoing example. Imagine, for instance, a game in which additional characters are introduced. These new characters look like the original (A) stimuli, but differ along some physical dimension (i.e. they are bigger, longer, or different in hue). Participant’s responses to such characters may be accounted for by the principle of stimulus generalisation. Stimulus generalisation has been defined as, “the spread of the effects of reinforcement (or other operations such as extinction or punishment) from one stimulus to other stimuli differing from the original along one or more dimensions” (Catania, 1998, p.391). As the new characters in the game share physical similarities to the original characters, we would expect the

appropriate responses to transfer to these new characters with little difficulty. However, there are not many modern games that rely on stimulus generalisations as part of game-play. Such a design is more typical of pioneering work in the early days of software engineering. Indeed, simpler psychological processes such as stimulus generalisation and simple conditioned reflexes may provide us with some understanding of the appeal of more rudimentary games such as *Pong*, *Pac-man* etc. (Loftus & Loftus, 1983). Nevertheless, stimulus generalisation is one simple process that may be involved in some aspects of computer games and is easily accounted for using a traditional behavioural principle.

Now imagine that even more characters are introduced to our simple game. These characters must be responded to, not based on their appearance but on their relationship to the original characters. Let us call the two original characters A1 and A2. Two new characters, labelled B1 and B2, are introduced as allies of A1 and A2, respectively. Now two further characters, C1 and C2, are introduced as allies of B1 and B2, respectively. These relationships can be established most easily in the game through instructions, as they often are in complex strategy games, or simply through interaction with the game (i.e., trial and error). The important point is that no relation between the A and C characters is explicitly specified at any stage. Nevertheless, most game players will have little difficulty responding to the C characters appropriately. This, in effect, is derived relational responding, or more specifically, the transfer of the response functions of the A stimuli to the C stimuli. Of course, real-world games involve the foregoing features wrapped in impressive presentation formats and a range of impressive graphics, but at a technical and psychological level we can conceive of at least some aspect of gaming as involving derived relational responding. Indeed, the purpose of this research is to test this very approach and its effectiveness as a paradigm for the psychological analysis of game playing.

In the current research, a series of games will be designed that will require participants to respond to both stimulus generalisations and derived relations in order to achieve a high score. This will test the suitability of a modern behavioural model of games and its utility as a conceptual framework for future research. Moreover, by introducing controlled simulated network delays we will be able to examine not only their effect on several subjective and objective features of game performance, but also how these delays interact with games of varying relational complexity. In Experiment 1, participants were first presented with Stimulus Equivalence Training and Testing with a range of stimuli to be used in a subsequent computer game. Participants then progressed to the Gaming Phase, in which they were required to play one of two games. Both games required participants to demonstrate stimulus generalisation and stimulus equivalence in order to solve the

problems posed in the game and to gain the highest possible score. Both games were essentially identical, except for the inclusion of delays in one game in order to simulate the Network Latency often experienced by on-line gamers. There were three levels to each game. Upon the completion of each level, a questionnaire was administered to measure participants' opinions of the preceding level using a range of subjective scales.

2.1.1 Method

2.1.1.1 Participants

Twenty undergraduate students (age range 18-25; ten male and ten female) were recruited for Experiment 2.1 through personal contacts.

2.1.1.2 Materials

The experiment was conducted in a quiet room free of distraction. All stimuli were presented on a computer screen (resolution 1024 x 768) via Microsoft Visual Basic 6.0 software (Cabello, Barnes-Holmes, O'Hora, & Stewart, 2002; Dixon & MacLin, 2003) which also recorded the nature and timing of stimulus presentation and participant responses. Eight stimuli were employed including two nonsense syllables, "JOM," (B1) and, "VEK," (B2) and six coloured geometric shapes (a red circle [A1], a red sphere [A1g], a green square [A2], a green cube [A2g], a yellow pentagon [C1] and a blue triangle [C2]).

2.1.1.3 Design

The experiment employed a 2X3 mixed between-within participants design with level of simulated delay as the between-participants variable and level of complexity as the within-participants variable. There were four dependent measures; participants' score on each level of the game, and their subjectively rated level of enjoyment, frustration and difficulty.

2.1.1.4 Procedure

Experiment 1 was divided into two stages; (i) the Stimulus Equivalence Training and Testing stage, and (ii) the gaming stage.

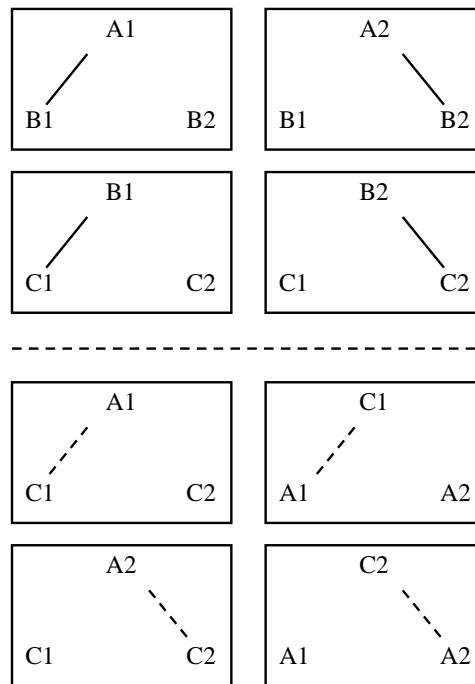


Figure 2.1: Matching-to-sample tasks for stimulus equivalence training (upper panel) and testing (lower panel).

(i) Stimulus Equivalence Training and Testing phase

The purpose of this first stage was to establish two three-member stimulus equivalence classes (A1-B1-C1 and A2-B2-C2) using a standard matching-to-sample (MTS) procedure. At the start of training, participants were seated comfortably in front of a standard 15" computer screen and keyboard before being instructed to choose which of two stimuli presented at the bottom of the computer screen (henceforth referred to as comparison stimuli) 'goes with' a sample stimulus presented in the middle of the screen. Participants' choices were guided by the corrective feedback provided following each response. A choice was registered when the participant clicked on one of the two comparison stimuli using the computer mouse. Trials were separated by an inter-trial interval (ITI) of 1 second. During testing for derived stimulus equivalence relations no feedback was provided and a choice was followed only by the ITI. Figure 2 presents a representation of the matching to sample tasks presented in the Stimulus Equivalence Training and Testing phase. The alphanumeric characters represent both the nonsense syllables and coloured shapes presented during the training and testing phases. Solid arrows represent trained relations and hashed arrows represent expected derived relations.

Training consisted of A-B and B-C training phases. On each trial of the A-B phase, either A1 or A2 appeared as sample while both the B stimuli appeared as comparisons at the bottom of the screen, with their left or right positions counter-balanced across trials to control for positional responding. In the presence of A1, choosing B1 was designated correct. Given A2 as a sample, choosing B2 was designated correct. On each trial of the B-C phase, either B1 or B2 appeared as sample while both the C stimuli appeared as comparisons at the bottom of the screen with their left or right positions counter-balanced across trials. In the presence of B1, choosing C1 was designated correct whilst in the presence of B2, choosing C2 was designated correct. Both phases were conducted in blocks of 20 trials, during which participants had to respond correctly on 19 out of 20 trials in order to pass that phase. The training on each phase continued indefinitely (or for a maximum period of 30 minutes) until the participant reached the specified response criterion. Tasks for each phase were presented in a quasi-random order with no more than two consecutive presentations of any task.

Following training, participants were exposed to derived equivalence testing. Testing involved four tasks (see Figure 1) presented 10 times each in a quasi-random order, with no more than two consecutive presentations of any task. These test tasks probed for both transitive A-C relations (i.e., A1-C1 and A2-C2) and combined transitive and symmetrical C-A relations (i.e., C1-A1 and C2-A2). Participants were required to respond correctly on 39 trials in a single block of 40 in order to pass testing. If they did not meet this criterion, they were returned to the beginning of the entire training and testing stage. Participants were cycled through training and testing until they passed the testing phase or until they failed four times in total. However, no participant required exposure to this stage more than twice and thus no participant was removed from the study at this stage. Participants typically required approximately fifteen minutes to complete the entire training and testing stage.

(ii) Gaming phase

Upon successful completion of equivalence testing, participants were exposed to a computer game consisting of three levels. Level 1 was a training level in which participants learned how to play a game that employed the A1 and A2 stimuli as game characters. Successful performance in Level 2 required stimulus generalisation to novel but physically similar game characters (A1g and A2g). Level 3 involved the use of the C stimuli in the place of the A stimuli. Success in Level 3 required the transfer of Level 1 response functions through stimulus equivalence class (i.e., a transfer of the response functions of A1 and A2 to C1 and C2, respectively).

The instructions for Level 1 were as follows:

In the following game you must earn as many points as possible in the time given, by destroying or saving the objects on screen. Objects can be saved by clicking on the object and destroyed through clicking on the destroy button. PLEASE NOTE THAT YOUR SCORE IS DISPLAYED IN THE BOTTOM RIGHT HAND CORNER OF THE SCREEN. The player who completes the entire game with the highest score will receive a prize. You must reach 20 points in order to finish this level. When you are ready, click 'Continue' to begin

As illustrated in Figure 2.2, the interface involved a control panel at the bottom of the screen, with the current game level displayed in the left hand corner, the participant's score presented in the right hand corner and a button labeled 'DESTROY!' in the centre of the panel. Stimuli were presented quasi-randomly on-screen in one of twelve possible locations, for a duration of 1.5s each. The area above the control panel wherein the stimuli were presented was called the 'game space.'

When a mouse click on a character was recorded, the character displayed was removed from the screen and the score displayed in the bottom right hand corner of the screen was adjusted accordingly. If the response was correct, the score was increased by 1. If the response was incorrect, or no response was recorded, the score was reduced by 1. When a click on the 'DESTROY!' button was recorded, the character displayed was removed from the screen and the button immediately became grey for 100 milliseconds before returning to its original appearance. The score displayed in the bottom right hand corner of the screen was adjusted accordingly. If the response was correct, the score was increased by 1. If the response was incorrect, or no response was recorded, the score was reduced by 1. If a participant failed to make either a 'save' or 'destroy' response within 1.5 seconds of the presentation of the character, that character was removed from the screen and the participant's score was reduced by one point. Regardless of the response made, a 1000 millisecond inter-trial interval was initiated after each response and before the presentation of the following trial.

Relational Complexity

In Level 1, hereafter referred to as the training level, the characters presented in the game space were the 'A' stimuli from the previous, equivalence training stage. In this level, points were earned for saving the A1 stimulus (a red circle) and destroying the A2 stimulus (a green square) within the 1.5 seconds given for a response. Points were lost for destroying the A1 stimulus, saving the A2 stimulus, or not responding within 1.5 seconds. There were twelve trial types, which corresponded to twelve possible positions the characters could appear on-screen. Seven of these trial types involved the presentation of the A1 stimulus, while five trial types involved the presentation of the A2

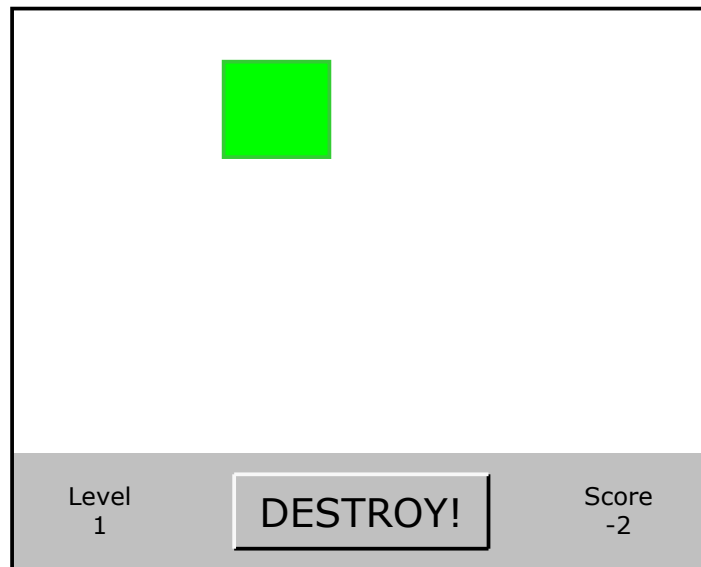


Figure 2.2: Representation of one task in the training phase of the game.

stimulus. Participants played this level until a score of 20 points had been achieved. It must be noted that there was no limit placed on the number of trials a participant could be exposed to during Level 1. Specifically, an incorrect response or a failure to respond within the given time limit resulted in a score of -1 for that trial. Thus, consistently correct responding was required for a participant to reach a score of 20.

The following instructions were presented before Levels 2 and 3 of the game:

In this level, you must score as many points as possible in the time given. YOUR SCORE WILL NOT BE DISPLAYED until you have finished the level. When you are ready, click the 'Continue' button to begin.

In Level 2, hereafter referred to as the generalisation level, the characters presented in the game space were 3-D representations of the A1 and A2 stimuli used in Level 1 (i.e., a green cube and a red sphere), hereafter referred to as A1g and A2g. Participants could gain a high score by responding to the 3-D objects in this level in the same manner in which they responded to the 2-D objects presented in Level 1. In Level 2 there were 48 trials in which the participant was required to score as many points as possible. As in Level 1, there were twelve trial types, which corresponded to twelve possible positions in which the characters could appear on-screen. Seven of these trial types involved the presentation of the A1g stimulus, while five trial types involved the presentation of the A2g stimulus. Scores were not displayed during game play, as the presence of a score indicator on a

trial-by-trial basis would have rendered Level 2 a training rather than a testing phase. That is, trial-by-trial feedback would quickly bring participants' responses under the direct contingency control of the displayed score. In effect, Level 2 would not function as a test for the generalisation of the contingencies established in Level 1 (i.e., 'save' A1 and 'destroy' A2) but simply reflect control by current contingencies during Level 2 (i.e., participants' ability to learn what was required to produce a score on each trial). This level ended once 48 trials had been completed. After the last trial in the level was completed, the participants' score for that level was presented in a box in the middle of the screen for a duration of 5 seconds.

In Level 3, hereafter referred to as the equivalence level, the characters that were presented in the game space were the 'C' stimuli from the equivalence training stage (i.e., C1 and C2). The same scoring system applied as in Levels 1 and 2. Points were gained for responding in accordance with the established equivalence relations. For example, as points were gained in Level 1 for saving the A1 stimulus (a red circle), points were gained in Level 3 for saving the C1 stimulus (a yellow pentagon), which participated in a derived equivalence relation with the red circle. Similarly, just as points were gained in Level 1 for destroying the A2 stimulus (a green square), points were gained in Level 3 for destroying the C2 stimulus (a blue triangle). As in Levels 1 and 2, there were twelve trial types, which corresponded to twelve possible positions in which the characters could appear on-screen. Seven of these trial types involved the presentation of the C1 stimulus, while five trial types involved the presentation of the C2 stimulus. There were 48 trials in which participants were required to score as many points as possible. Upon the completion of the 48 trials, Level 3 came to an end and the participant's score was displayed in a box in the middle of the screen for a duration of 5 seconds.

Modelling Network Delay

Before commencing the game, participants were randomly assigned to one of two conditions - 'Delay' or 'Non-Delay'. In both conditions, the game played was as described above. However, in the Delay condition, the game included simulated network delays while in the other condition it did not. The delays were designed to emulate a game network delay in which a response made by a player is not registered with the game server within the necessary amount of time. Simulated network delays were inserted on three out of every five trials on which A2, A2g and C2 were presented (hereafter referred to as stimulus equivalence class 2). Delays were implemented by "freezing" the 'destroy' button for the duration of a trial. More specifically, as soon as the 'destroy' button was pressed, it became grey and rather than returning to its original appearance after 100 milliseconds, remained grey for the remainder of the trial.

Additionally, the character (i.e., A2, A2g or C2, depending on the Level being played) remained on-screen and the score did not change for the remainder of the trial. Any further responses made during the time allotted for that particular trial, either using the ‘destroy’ button, or by clicking on the character itself, were disabled. When 1.5 seconds had elapsed, the character was removed from the screen and a point was deducted from the participants’ score, regardless of responses made during the delay period. Thus, it appeared to the participant that their response had been ineffective and the time given for a response had simply run out without a response being registered.

When a participant responded correctly on a delay trial, a point was deducted from the score displayed. As such, Delay condition participants necessarily achieved lower scores on the display than those in the Non-Delay condition. However, the display score was not the same as the ‘total correct’ score recorded by the computer software and used in the current data analysis. More specifically, the number of correct first responses made by a participant to the trials presented in a given level were recorded without the participants’ awareness. Thus, Correct Responses made during a ‘delay’ trial, while leading to a loss of points on the score displayed on-screen to participants, were actually recorded as Correct Responses for the purposes of data analysis. In this way, it was possible for all participants in all conditions to achieve a score of 100%. In effect, this method provided a ‘total correct’ score that was not confounded by the actual physical limiting effects of delay. The ‘total correct’ measure is an index of the psychological effect of delay on participants’ ability to produce Correct Responses during the game.

Upon the completion of all three levels of the game, a brief questionnaire was presented to participants on-screen. Instructions presented at the top of the screen were as follows:

“Congratulations! You have reached the end of this level. Please answer the following questions by moving the sliders to your desired positions and then clicking submit.”

Three questions were presented on-screen simultaneously. These questions were identical to those used by Prendergier, Mori and Ishizuka (2005) in their study of user frustration in computer game playing. Participants indicated their choice by sliding a bar to their preferred position on a ten-point Likert scale presented on the computer screen. The questions posed were: (i) Did you find the game difficult?; (ii) Did you enjoy playing the game?; (iii) Are you frustrated with the game? There was no time limit imposed for responding to the three questions, which were presented on-screen simultaneously. Once all questions were answered, the participant clicked a ‘Submit’ button and their responses were recorded by the computer. At this point, their participation in the experiment was complete and they were thanked and debriefed.

Table 2.1: Number of attempts required by each participant to pass each stage of Stimulus Equivalence Training and Testing

	Training 1	Training 2	Testing	Training 1	Training 2	Testing
Non-Delay Game Participants						
Participant 1	2	1	1			
Participant 2	1	1	1			
Participant 3	2	1	1			
Participant 4	1	1	1			
Participant 5	1	1	x	1	1	1
Participant 6	2	2	1			
Participant 7	2	1	1			
Participant 8	2	1	1			
Participant 9	2	2	1			
Participant 10	1	2	x	1	1	1
	Training 1	Training 2	Testing	Training 1	Training 2	Testing
Delay Game Participants						
Participant 11	1	1	1			
Participant 12	1	1	1			
Participant 13	1	1	x	1	1	1
Participant 14	1	2	1			
Participant 15	1	2	1			
Participant 16	1	1	1			
Participant 17	1	3	x	1	1	1
Participant 18	1	1	x	1	1	1
Participant 19	1	1	1			
Participant 20	1	1	1			

2.1.2 Results and Discussion

2.1.2.1 Stimulus Equivalence Training and Testing

All participants passed Stimulus Equivalence Training and Testing within two exposures to the training and testing procedure, with fifteen passing on their first exposure. All participants completed this stage of the experiment within twenty minutes. Table 2.1 presents the number of attempts each participant required to pass each stage in Stimulus Equivalence Training and Testing. Note that where an ‘x’ appears under testing, this means that the participant did not pass the equivalence test and required further training.

2.1.2.2 Game Playing

Network Delay

All 20 participants passed the Stimulus Equivalence Training and Testing procedure and advanced to the gaming stage. Total Correct Responses are not presented for the Training Level, as

there was a large variation in the number of trials presented across participants. The reason for this was that it was possible for participants to lose as well as gain points during this phase (see Method section 2.1.1.4). As a result, the number of trials taken to achieve the passing criterion score of +20 varied considerably across participants. Thus, Total Correct Responses are not a useful indicator of performance in the Training Level.

Ratings for Enjoyment, Difficulty and Frustration were collected through a questionnaire administered after each level of the game. Participants were required to rate whether they agreed or disagreed with a number of statements about the previous level of the game, using a 10 point Likert scale. Table 2.2 displays mean ratings for Enjoyment, Difficulty and Frustration for the Training Level.

Training Level

Table 2.2: Mean scores for all measures employed on the Training Level of both the Non-Delay (n=10) and Delay (n=10) Games.

	Non-Delay	Delay
Correct Responses	N/A	N/A
Enjoyment	6.4	6.3
Frustration	1.9	6.4
Difficulty	2.4	5.9

Mean Enjoyment ratings are relatively consistent across both game types in the training level. This suggests that participants preferred neither the delay nor the Non-Delay Games at the training level. However, there is a large difference in frustration ratings across the two game types. Participants appeared to rate the Delay Game as more frustrating than the Non-Delay Game. Additionally, there is a large difference in difficulty ratings across game types.

Generalisation Level

Correct Responses were calculated as the total number of correct first responses to trials in a given level. It is important to remember that there is a distinction between the score displayed to participants during game play and the total number of Correct Responses made (see Procedure section 2.1.1.4). Only the latter is used in the following analysis. Table 2.3 presents the number of Correct Responses made by each participant in the generalisation level.

A Correct Response rate of 43 Correct Responses on a block of 48 trials (or 89.583%, hereafter referred to as a 90% correct criterion) was deemed a pass score for Levels 2 and 3. If a participant

Table 2.3: Number of Correct Responses made by each participant in the Generalisation Level

No Delay									
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
48	48	46	48	47	36	46	47	48	31
Delay									
P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
31	40	28	35	47	47	46	46	38	42

failed to reach this criterion on their first exposure to a level, performance for that level was recorded as a fail. Eight out of a total of ten participants who played the Non-Delay Game passed the generalisation level at the 90% correct criterion. Four out of a total of ten participants who played the Delay Game passed the generalisation level at the 90% correct criterion.

Table 2.4: Mean scores for all measures employed on the Generalisation Level of both the Non-Delay (n=10) and Delay (n=10) Games.

	Non-Delay	Delay
Correct Responses	44.5	40
Enjoyment	6.2	5.2
Frustration	3.6	7.2
Difficulty	3	6.2

Table 2.4 displays the mean Correct Responses recorded during game play and mean ratings for Enjoyment, Difficulty and Frustration for the Generalisation Level. Mean Correct Responses are higher for the Non-Delay Game than the Delay Game. Enjoyment ratings appear to be higher for the Non-Delay Game than the Delay Game, suggesting that participants preferred the game that was free of delays. Indeed, participants also rated the Delay Game as more frustrating than the Non-Delay Game. Additionally, ratings of difficulty were higher for the Delay Game.

Equivalence Level

Table 2.5: Number of Correct Responses made by each participant in the Equivalence Level

No Delay									
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
46	47	45	46	48	45	48	48	47	45
Delay									
P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
16	38	28	35	48	43	44	45	41	47

Table 2.5 presents the number of Correct Responses made by each participant in the equivalence level. All ten participants who played the Non-Delay Game passed the generalisation

level at the 90% correct criterion. Five out of a total of ten participants who played the Delay Game passed the generalisation level at the 90% correct criterion.

Table 2.6: Mean scores for all measures employed on the Equivalence Level of both the Non-Delay (n=10) and Delay (n=10) Games.

	Non-Delay	Delay
Correct Responses	46.5	38.5
Enjoyment	6.8	5.8
Frustration	3.7	6.5
Difficulty	2.4	6.9

Table 2.6 displays mean Correct Responses recorded during game play and mean ratings for Enjoyment, Difficulty and Frustration for the Equivalence Level. Mean Correct Responses are higher for the Non-Delay Game than the Delay Game. Enjoyment ratings are higher for the Non-Delay Game than the Delay Game, suggesting that participants preferred the game that was free of delays. Moreover, participants also rated the Delay Game as more frustrating than the Non-Delay Game. In addition, ratings of difficulty were higher for the Delay Game.

Relational Complexity

Table 2.7: Mean scores for all measures employed during all three levels of both the Non-Delay (n=10) and Delay (n=10) Games.

	Training	Generalisation	Equivalence
Non-Delay Game			
Correct Responses	N/A	44.5	46.5
Enjoyment	6.4	6.2	6.8
Frustration	1.9	3.6	3.7
Difficulty	2.4	3	2.4
Delay Game			
Correct Responses	N/A	40	38.5
Enjoyment	6.3	5.2	5.8
Frustration	6.4	7.2	6.5
Difficulty	5.9	6.2	6.9

Table 2.7 displays mean Correct Responses recorded during game play and mean ratings for Enjoyment, Difficulty and Frustration for all three levels of the game. Mean Correct Responses increased from Level 2 to Level 3 of the Non-Delay Game, whereas they decreased in the game that contained delays. Mean Enjoyment ratings were relatively stable across all three levels of both game types, suggesting that relational complexity, as modelled in the current study, had little impact on participants' enjoyment of those games. Apart from a large increase from the training to

generalisation level in the Non-Delay Game, mean ratings of frustration were also relatively consistent across levels in both games. Difficulty ratings were consistent across all three levels in the Non-Delay Game, while a gradual increase in mean difficulty ratings was observed in the game containing delays. This finding may suggest some interaction between the variables of relational complexity and network delay.

Table 2.8: Results of a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Correct Responses, difficulty, enjoyment and frustration (n=20).

	Wilks' Lambda	F Value	p Value	Eta Squared
Correct Responses				
Complexity	0.998	0.044	0.837	0.002
Delay	N/A	5.039	0.038*	0.219
Interaction	0.894	2.135	0.161	0.106
Enjoyment				
Complexity	0.866	1.310	0.296	0.134
Delay	N/A	0.354	0.559	0.019
Interaction	0.874	1.224	0.319	0.126
Frustration				
Complexity	0.879	1.172	0.334	0.121
Delay	N/A	16.489	<0.01*	0.478
Interaction	0.927	0.668	0.526	0.073
Difficulty				
Complexity	0.938	0.561	0.581	0.062
Delay	N/A	21.157	<0.01*	0.54
Interaction	0.901	0.932	0.413	0.099

Table 2.8 presents the results of a mixed between-within subjects analysis of variance, conducted to explore the impact of both Relational Complexity and Level of Delay on participants' Correct Responses and self-reports of Difficulty, Frustration and Enjoyment. Note that p-values marked with an asterisk represent significant results. There was a significant effect for Level of Delay on Correct Responses, and ratings of Difficulty and Frustration. Level of Delay had no significant effect on Enjoyment ratings. Relational Complexity had no significant effect on any measure employed in the study. In addition, no significant interaction effects were observed between the two variables of interest on any measure.

These data suggest that simulated network delay had a significant effect on the number of Correct Responses made during the computer games. Moreover, these delays influenced participants' self-reported levels of Difficulty and Frustration. However, delay was not found to have had a significant impact on self-reported levels of Enjoyment. Furthermore, Relational Complexity also failed to impact on participants' Correct Responses or any of the subjective measures. Thus, we can conclude that the presence of network delays in the current study was detrimental to the game

playing experience even though there was no significant difference in Enjoyment ratings across the two game types.

One possible criticism of the current study that may qualify these findings is that simulated network delays were applied to stimuli from one stimulus equivalence class only (i.e., A2-A2g-C2). However, the application of network delays to one stimulus class only lead to an unforeseen problem with identifying the source of differences found across game types. That is, as the delay was only applied to stimuli from the same class, it was possible to complete the game by ignoring these stimuli and responding only to the stimuli presented without delay (i.e., A1, A1g and C1). Some participants may have adopted this strategy and used it in subsequent levels of the game, even though this strategy would impact negatively on the number of Correct Responses. In effect, it is possible that some participants might have completed all three levels of the game successfully without fully experiencing the simulated network delays. This possibility may help account for the differences in the number of Correct Responses between the delay and Non-Delay Games. Experiment 2 was conducted in order to address this issue.

2.2 Experiment 2.2

The results of Experiment 2.1 showed that simulated network delay had a significant effect on the number of Correct Responses made during the computer game and on the self-reported subjective measures of Difficulty and Frustration. However, it is possible that some participants completed the game by ignoring the stimuli that were presented with simulated delays (i.e., A2, A2g and C2). A participant adopting this strategy could therefore play the game successfully without fully experiencing the simulated network delays. In order to address this issue, Experiment 1 was replicated with a modified method of simulated delay. Specifically, in Experiment 2, delays were dispersed across the entire range of stimuli employed. As in Experiment 1, each level involved the presentation of twelve trial types, five of which involved stimuli from equivalence class 2 (i.e., A2, A2g or C2, depending on the Level) and seven of which involved stimuli from class 1 (i.e., A1, A1g, or C1, depending on the Level). Simulated network delays were inserted on three of the twelve trial types in each level. Half of those participants who played the Delay Game received a game containing delays on 2/7 of equivalence class 1 trials and 1/5 of class 2 trials, while the remaining half played a game in which there were delays on 1/7 of class 1 trials and 2/5 of class 2 stimuli. Groups were balanced for gender.

2.2.1 Method

2.2.1.1 Participants

Ten volunteers (age range 18-25; five male and five female) were recruited for Experiment 2.2.

2.2.1.2 Materials

The materials used were the same as in Experiment 2.1.

2.2.1.3 Design

The ten participants recruited for Experiment 2.2 were all assigned to the Delay Game condition. The performances of the ten Non-Delay Game participants from Experiment 2.1 were retained for comparison during data analysis. Thus, Experiment 2.2 mixed between-within participants design, with level of simulated delay as the between-participants variable and level of complexity as the within-participants variable. There were four dependent measures; participants' score on each level of the game, and their subjectively rated level of enjoyment, frustration and difficulty.

2.2.1.4 Procedure

Participants were first exposed to Stimulus Equivalence Training and Testing identical to that used in Experiment 2.1. Once Stimulus Equivalence Training and Testing was passed, participants were presented with the Delay Game. The Delay Game in this study was identical to that in Experiment 2.1, with the exception of the method in which the delay was implemented. Specifically, in Experiment 2.2, delays were spread across both stimulus sets. That is, each level involved the presentation of twelve trial types, five of which involved a stimulus from class 2 (i.e., A2, A2g or C2) and seven of which involved a stimulus from class 1 (i.e., A1, A1g or C1). Simulated network delays were inserted on three of the twelve trial types in each level. Half of those participants who played the Delay Game received a game containing delays on 2/7 of equivalence class 1 trials and 1/5 of class 2 trials, while the remaining half played a game in which there were delays on 2/7 of class 1 trials and 1/5 of class 2 stimuli. Groups were balanced for gender.

A delay on a trial involving the presentation of equivalence class 2 members was implemented exactly as described in the procedure section for Experiment 2.1. However, as equivalence class 1 trials required a different response than equivalence class 2 trials, class 1 trials also required a different method of modelling delay. Specifically, A1, A1g and C1 were presented in one of the seven

possible locations in the game space, depending on which level was being played. At first, trials involving these stimuli looked no different from a fully functional Non-Delay trial. However, once the character was clicked on with the mouse, it was removed from the screen for 50 milliseconds before returning for the remainder of the trial. This produced a ‘flicker’ effect, designed to signal that the response had been ineffectual. Once the stimulus re-appeared on the screen, it remained there for the rest of the 1.5s trial and could not be removed through further clicking on the object or on the destroy button. Additionally, the score did not change for the remainder of the trial. When 1.5s had elapsed, the character was removed from the screen and a point was deducted from the participant’s score. Thus, it appeared to the participant that their response had been ineffective and the time given for a response had simply run out without a response being registered.

2.2.2 Results and Discussion

2.2.2.1 Stimulus Equivalence Training and Testing

All participants passed Stimulus Equivalence Training and Testing within two exposures to the training and testing procedure, with thirteen passing on their first exposure. All participants completed this stage of the experiment within twenty minutes. Table 2.9 presents the number of attempts each participant required to pass each stage in Stimulus Equivalence Training and Testing. Note that where an ‘x’ appears under testing, this means that the participant did not pass the equivalence test and required further training.

2.2.2.2 Game playing

(i) Network Delay

All Participants passed the Stimulus Equivalence training and Testing phase and advanced to the Game phase. A Training level was initially presented, in which participants learned how to play a game that employed the A1 and A2 stimuli as game characters. As in Experiment 2.1, Correct Responses are not presented for this level.

Training Level

Table 2.10 displays mean ratings for Enjoyment, Difficulty and Frustration for the Training Level. Mean Enjoyment ratings are consistent across both game types in the training level. This suggests that participants preferred neither the delay nor the Non-Delay Games at the training level. However, there is a large difference in frustration ratings across the two game types. Participants

Table 2.9: Number of exposures required by all participants to pass each stage of Stimulus Equivalence Training and Testing phase

	Training 1	Training 2	Testing	Training 1	Training 2	Testing
Non-Delay Game Participants						
Participant 1	2	1	1			
Participant 2	1	1	1			
Participant 3	2	1	1			
Participant 4	1	1	1			
Participant 5	1	1	x	1	1	1
Participant 6	2	2	1			
Participant 7	2	1	1			
Participant 8	2	1	1			
Participant 9	2	2	1			
Participant 10	1	2	x	1	1	1
	Training 1	Training 2	Testing	Training 1	Training 2	Testing
Delay Game Participants						
Participant 11	3	1	x	1	1	1
Participant 12	1	1	1			
Participant 13	1	1	1			
Participant 14	2	1	x	1	1	1
Participant 15	2	1	x	1	1	1
Participant 16	1	1	1			
Participant 17	1	1	x	1	1	1
Participant 18	4	1	x	1	1	1
Participant 19	1	1	1			
Participant 20	1	2	1			

appeared to rate the Delay Game as more frustrating than the Non-Delay Game. Additionally, there is a noticeable difference in difficulty ratings across game types.

Generalisation Level

Correct Responses were calculated as the total number of correct first responses to trials in a given level. Table 2.11 presents the number of Correct Responses made by each participant in the generalisation level.

Eight out of a total of ten participants who played the Non-Delay Game passed the generalisation level at the 90% correct criterion. Seven out of a total of ten participants who played the Delay Game passed the generalisation level at the 90% correct criterion.

Table 2.12 displays mean Correct Responses recorded during game play and mean ratings for Enjoyment, Difficulty and Frustration for the Generalisation Level. Mean Correct Responses are relatively consistent for both the Non-Delay and Delay Games. Enjoyment ratings also appear to be consistent across both game types, suggesting that participants did not prefer one game over the

Table 2.10: Mean scores for all measures employed on the Training Level of both the Non-Delay (n=10) and Delay (n=10) Games.

	Non-Delay	Delay
Correct Responses	N/A	N/A
Enjoyment	6.4	6.5
Frustration	1.9	5.1
Difficulty	2.4	4.9

Table 2.11: Number of Correct Responses made by each participant in the Generalisation Level

No Delay									
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
48	48	46	48	47	36	46	47	48	31
Delay									
P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
40	31	45	47	48	39	47	47	48	46

other. Participants rated the Delay Game as more frustrating than the Non-Delay Game. Ratings of difficulty were also higher for the Delay Game.

Equivalence Level

Table 2.13 presents the number of Correct Responses made by each participant in the Equivalence Level. All ten participants who played the Non-Delay Game passed the generalisation level at the 90% correct criterion. Three out of a total of ten participants who played the Delay Game passed the generalisation level at the 90% correct criterion.

Table 2.14 displays mean Correct Responses recorded during game play and mean ratings for Enjoyment, Difficulty and Frustration for the Equivalence Level. Mean Correct Responses are higher for the Non-Delay Game than the Delay Game. Enjoyment ratings appear to be consistent across both game types. Participants rated the Delay Game as marginally more frustrating than the Non-Delay Game. In addition, ratings of difficulty were higher for the Delay Game.

(ii) Relational Complexity

Table 2.15 displays mean Correct Responses recorded during game play and mean ratings for Enjoyment, Difficulty and Frustration for all three levels of the game. Mean Correct Responses increase from Level 2 to Level 3 of the Non-Delay Game, whereas they decrease in the game which contained delays. Mean Enjoyment ratings are relatively stable across all three levels of both game types, suggesting that relational complexity, as modelled in the current study, had little impact on

Table 2.12: Mean scores for all measures employed on the Generalisation Level of both the Non-Delay (n=10) and Delay (n=10) Games.

	Non-Delay	Delay
Correct Responses	44.5	43.8
Enjoyment	6.2	6.4
Frustration	3.6	5.1
Difficulty	3	5.1

Table 2.13: Number of Correct Responses made by each participant in the Equivalence Level

No Delay									
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
46	47	45	46	48	45	48	48	47	45
Delay									
P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
38	37	33	47	31	41	47	5	39	47

participants' enjoyment of those games. Apart from a large increase from the training to generalisation level in the Non-Delay Game, mean ratings of frustration are also relatively consistent across levels in both games. Difficulty ratings are consistent across all three levels in the Non-Delay Game, while a gradual increase in mean difficulty ratings is observed in the game containing delays. This finding may suggest some interaction between the variables of relational complexity and network delay.

Table 2.8 presents the results of a mixed between-within subjects analysis of variance, conducted to explore the impact of both Relational Complexity and Level of Delay on participants' Correct Responses and self-reports of Difficulty, Frustration and Enjoyment. Note that those p-values marked with an asterisk represent significant results. There was a significant effect for Level of Delay on Correct Responses, and ratings of Difficulty and Frustration. Level of Delay had no significant effect on Enjoyment ratings. Relational Complexity had no significant effect on any measure employed in the study. In addition, no significant interaction effects were observed between the two variables of interest on any measure. These results mirror those of Experiment 2.1, and thus support the findings of that study.

In conclusion, participants who played the game containing simulated network delays produced lower scores on the measure of Total Correct Responses than those who played the Non-Delay Game. In addition, those who played the Delay Game reported higher Difficulty and Frustration ratings than those who played the Non-Delay Game. Thus, in the game designed for the current study, the presence of simulated network delays significantly reduced participants' abilities to complete the game successfully, as well as significantly raising their perceived Difficulty of, and

Table 2.14: Mean scores for all measures employed on the Equivalence Level of both the Non-Delay (n=10) and Delay (n=10) Games.

	Non-Delay	Delay
Correct Responses	46.5	36.5
Enjoyment	6.8	6.3
Frustration	3.7	4.5
Difficulty	2.4	6.1

Table 2.15: Mean scores for all measures employed during all three levels of both the Non-Delay (n=10) and Delay (n=10) Games.

	Training	Generalisation	Equivalence
Non-Delay Game			
Correct Responses	N/A	44.5	6.5
Enjoyment	6.4	6.2	6.8
Frustration	1.9	3.6	3.7
Difficulty	2.4	3	2.4
Delay Game			
Correct Responses	N/A	43.8	36.5
Enjoyment	6.5	6.4	6.3
Frustration	5.1	5.1	4.5
Difficulty	4.9	5.1	6.1

Frustration with that game. It would appear that the presence of network delays in the current study were detrimental to the game playing experience. However, there was no significant difference in Enjoyment ratings across the two game types.

The current findings also suggest that the increased complexity of successive levels of the game did not significantly alter participants' performance on that game, nor their ratings of Enjoyment, Difficulty or Frustration. Thus, it would appear that Relational Complexity did not have a significant effect on participants' behaviour or experience in the current study.

2.3 Experiment 2.3

The previous two experiments suggest some role for a Relational Frame analysis in a psychological investigation into game playing, by providing a framework for understanding how people play computer games and precisely what makes them enjoyable. The study demonstrated that players can engage in relational activities as part of a game, and that a relatively popular game format can be understood and analysed in relational terms. Players demonstrated both stimulus generalisation and stimulus equivalence in a game playing environment without explicit instructions to do so.

Table 2.16: Results of a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Correct Responses, difficulty, enjoyment and frustration (n=20)

	Wilks' Lambda	F Value	p Value	Eta Squared
Correct Responses				
Complexity	0.936	1.237	0.281	0.064
Delay	N/A	5.252	0.034*	0.226
Interaction	0.825	3.808	0.067	0.175
Enjoyment				
Complexity	0.989	0.096	0.909	0.011
Delay	N/A	<0.01	0.944	<0.01
Interaction	0.974	0.228	0.799	0.026
Frustration				
Complexity	0.932	0.62	0.55	0.068
Delay	N/A	5.239	0.034*	0.225
Interaction	0.856	1.431	0.266	0.144
Difficulty				
Complexity	0.940	0.547	0.589	0.06
Delay	N/A	13.178	<0.01*	0.423
Interaction	0.919	0.753	0.486	0.081

In Level 2 of both experiments, players gained high scores for producing generalised response patterns acquired during the training level, in the presence of physically similar stimuli. Players gained high scores during Level 3 by demonstrating a transfer of the appropriate response function, acquired in Level 1, through derived equivalence relations. Indeed, in Experiment 2.1, 12/20 participants passed at a 90% correct response criterion in the stimulus generalisation level, while 15/20 passed at a 90% correct response criterion in the stimulus equivalence level. In Experiment 2.2, 15/20 participants passed at a 90% correct response criterion in the stimulus generalisation level, while 13/20 passed at a 90% correct response criterion in the stimulus equivalence level. Of those who did not reach the 90% correct response criterion, the majority performed at above chance levels in both the stimulus generalisation and stimulus equivalence levels, in a context where there are significant response time demands. Thus, if the response window for each trial in the previous study was infinite we would expect to see response accuracies approach 100%, given the performances observed. The important point however, is that a large number of participants were able to consistently respond in accordance with the trained equivalence relations. This finding provides support for the view that game players can engage in derived relational responding while playing games. Indeed, this demonstration may serve as a suitable analogue of some types of complex skills required in many commercially available games.

Experiments 2.1 and 2.2 found that Network Latency had a significant effect on Total Correct Responses, Difficulty and Frustration. Participants found the game containing simulated network

delays significantly more difficult and frustrating than the game which did not contain delays, and achieved significantly lower scores when playing the game containing delays. These effects were replicated across both experiments, suggesting that these findings are reliable and easily approached and analysed using the current derived relational procedures. Thus, these first two experiments represent a solid first step in understanding functionally the relationship between Network Latency and online game playing experience and performance.

Interestingly, neither of the first two experiments found that Network Latency had a significant effect on game enjoyment. One might point to the replicated failure to find an effect for Latency on enjoyment as suggesting that such an effect may be difficult or impossible to establish. However, given the face validity of the Latency variable as being detrimental to game enjoyment, other possibilities must be considered. It may be, for instance, that a clearer relationship between Latency and enjoyment was occluded to some extent by the non-standardised rating scales. Thus, what is required, in addition to further manipulation of the latency variable, is a more reliable subjective measure of enjoyment.

Unfortunately there is no standardised measure of enjoyment of computer games available in the literature. However, one scale developed by Kahneman, Krueger, Schkade, Schwartz and Stone (2004) as part of their Day Reconstruction Method (DRM) of measuring quality of life, may be of use given the foregoing concern. The particular scale in question asks participants to rate individual recent experiences for enjoyability using a short 12 item questionnaire. Results of a number of questions are then summed in order to give two constructs; Positive Affect and Negative Affect. Importantly, the DRM has been validated with a sample of 1018, (Kahneman et al, 2004) so the scale in question may provide us with a more reliable subjective measure than that employed in the previous study.

The current study will employ a new scale in order to more reliably measure participants' subjective experience of the games developed. A short 12-tem questionnaire, taken from Kahneman et al. (2004) will be employed in order to obtain ratings of Positive Affect and Negative Affect for all levels in the games played by participants. The current study also aims to extend the findings of the previous study by increasing the network latency variable up to a duration of 1s. This strategy is in line with findings from the field of engineering (i.e., Delaney, Meenaghan, Ward, & McLoone, 2004; Vaghi, Greenhalgh, & Benford, 1999), that suggest that incrementally increasing the level of delay in a game leads to a progressively less 'playable' game. Thus, the current study will model network latency at two levels, 0.5s and 1s, in order to determine whether or not there exists a functional relationship between the level of on-line delay in a game and performance and positive affect.

Perhaps the most interesting finding of Experiments 2.1 and 2.2 was that neither experiment found a significant effect for Relational Complexity on game performance or on any of the subjective measures. Relational Complexity was found not to influence significantly any of the measures employed across both experiments, nor to interact significantly with Network Latency in either experiment. As complexity, or challenge, has been identified in the games literature as being a very important factor in the experience and enjoyment of game playing (Gingold, 2005; Malone, 1982; Morlock, Yando, & Nigolean, 1985; Vorderer, Hartmann, & Klimmt, 2003) it would appear that the relationship between relational complexity and game enjoyment is not as clear cut as was originally envisaged. At the very least, it could be suggested that relational complexity was not sufficiently manipulated in the previous study to illustrate its relationship to enjoyment. Of course, further research into the effects of very high levels of complexity on subjective experience is required. For instance, research should involve increasing the complexity of the relational activities involved in a game by manipulating the relations between the relevant stimuli, or by removing the stimulus equivalence testing phase from the procedure. These strategies should allow greater experimental control over relational complexity and its effects on game playing.

The current study will attempt to manipulate relational complexity more effectively in order to examine more closely the effects of both Network Latency and Relational Complexity, as well as the relationship between them. Specifically, Nodal Distance has been identified as a variable which may affect performances in the formation of equivalence classes (Fields, Adams, Verhave, & Newman, 1990; Fields, Reeve, Rosen, Varelas, & Adams, 1997) and is a variable which should lend itself to manipulation along a continuum ranging from very low to very high complexity. Nodal Distance may be best explained in a situation where a large number of stimuli are trained sequentially in order to form an equivalence class. For example, imagine a preparation in which a stimulus A1 is matched with B1, then B1 is matched with C1, and so on, until a three-node relation has been formed (i.e., A1-B1-C1-D1-E1). A test for derived transitive relations between C1 and A1 can then be described as a 1-node derived relation, because one stimulus separates the A stimulus from the C stimulus in the linear relation between them. However, the derived relation between the A stimulus and the E stimulus is a three-node relation, because three stimuli separate A1 and E1.

Importantly, Nodal Distance has been identified as a key determinant of performance on equivalence tests (Fields et al., 1997). Specifically, participants have more difficulty forming equivalence classes involving a larger number of nodes, as measured by both accuracy and response times. As such, the current study will manipulate the nodal distance between stimuli presented in the game stage, through using a blocked sequential training procedure. Participants will be

presented with game characters that participate in a one-node equivalence relation with the stimuli used in a training level. In effect, this procedure is similar to that used in the Experiments 2.1 and 2.2. However, participants will also be exposed to a further game involving a three-node relation. We would expect to see lower scores on this latter game due to the requirement to form three-node relations in order to respond effectively. Because the three-node game represents more of a challenge than the one-node game, we would also expect to see higher ratings of positive affect and lower ratings of negative affect for this more complex game.

To examine these ideas, a series of games were designed in which participants were required to demonstrate derived relational responding in order to achieve a high score. Relational complexity and Network Latency were manipulated in order to examine their effects on both objective and subjective measures of game playing. Participants were first presented with Stimulus Equivalence Training and Testing. Once participants passed the Stimulus Equivalence Training and Testing phase they were presented with one of two game types, which are essentially identical, except for the length of simulated network delays inserted. One game contained 0.5s delays, while the other contained 1s delays. All games consisted of three levels; a baseline training level, a one-node stimulus equivalence level and a three-node stimulus equivalence level. Participants were required to produce robust equivalence responding in order to gain high scores in these games. In addition a new questionnaire was administered after each level in order to measure participants' subjective experiences.

2.3.1 Method

2.3.1.1 Participants

Twenty three participants were recruited, all of whom were undergraduate students. Twelve of these were male, while 11 were female. Participants were offered a payment of €5 upon completion of the game.

2.3.1.2 Materials

The experiment was conducted in a quiet room free of distractions. All responses and stimulus presentations were controlled by Microsoft Visual Basic 6.0 software (c.f., Cabello, Barnes-Holmes, O'Hora, & Stewart, 2002; Dixon & MacLin, 2003) and were presented on a computer screen (resolution 1024 x 768). Two nonsense syllables, "JOM," and, "VEK," and eight coloured geometric shapes (a red circle, green square, yellow pentagon, blue triangle, black crescent, white cylinder, cyan cross and orange arrow) were used as stimuli.

2.3.1.3 Design

The experiment employed a 2X2 mixed between-within participants design. The main variables were the length of simulated delay in each game, and the level of complexity across the three levels of the game. The first variable was manipulated across participant groups whereas the second variable was manipulated within groups (i.e., all participants were exposed to all three levels of the game). There were three dependent measures; participants' score on each level of the game, and their subjectively rated level of positive affect, negative affect, competence and impatience.

2.3.1.4 Procedure

The study was divided into two stages: the Stimulus Equivalence Training and Testing Phase, and the Game Phase.

Stimulus Equivalence Training and Testing

Participants received the Stimulus Equivalence Training and Testing stage first. During this stage, two three-node stimulus classes (A1-B1-C1-D1-E1 and A2-B2-C2-D2-E2) were established using the standard matching-to-sample procedure. The procedure employed was exactly as that employed in Experiments 2.1 and 2.2, apart from the fact that a further two relations were trained. Stimulus Equivalence Training was divided into four phases, the first in which 'A' stimuli (a red circle and a green square), were trained to 'B' stimuli (three-letter nonsense syllables; JOM and VEK); the second in which 'B' stimuli were trained to 'C' stimuli (a blue triangle and a yellow pentagon); the third in which 'C' stimuli were trained to 'D' stimuli shapes (a black crescent and a white cylinder); and the fourth in which 'D' stimuli were trained to 'E' stimuli (a cyan cross and an orange arrow).

Once participants had passed Stimulus Equivalence Training they were presented with Stimulus Equivalence Testing. The Stimulus Equivalence Testing phase involved four tasks presented 10 times each in a quasi-random order, with no more than two consecutive presentations of any task. The test tasks probed for both transitive A-E relations (i.e., A1-E1 and A2-E2) and combined transitive and symmetrical E-A relations (i.e., E1-A1 and E2-A2). Participants were required to respond correctly on 39 trials in a single block of 40 in order to pass this phase. If they did not meet this criterion, participants were returned to the beginning of the Stimulus Equivalence Training and Testing sequence (up to a maximum of three times) until they once again passed the training phase and then the testing phase on their first exposure. Participants typically required approximately twenty to twenty five minutes to complete the entire equivalence training and testing stage.

Gaming

Upon completion of the Stimulus Equivalence Training and Testing procedure, participants were assigned to play one of two games, a Short Delay Game or a Long Delay Game. Both games were essentially identical, with the difference that one game included 0.5s simulated network delays while the other contained 1s delays. Both games consisted of three levels; a training level, a one-node stimulus equivalence level and a three-node equivalence level.

Each level involved the quasi-random presentation of twenty four trial types, twelve of which involved the presentation of stimulus class 1 stimuli (i.e., A1, C1, or E1), and twelve of which involved the presentation of the class 2 stimuli (i.e., A2, C2, or E2, depending on the level). All levels in both games featured an identical user interface to that used in Experiment 2.1. Unlike Experiment 2.1, the stimuli presented in Experiment 2.3 increased in size by 25% every 0.5 of a second from the onset of the stimulus presentation, in order to simulate the effect of the stimulus approaching the player in three-dimensional space. The following instructions were presented as text on screen at the beginning of Level 1 of the game:

In the following game you must earn as many points as possible in the time given, by destroying or saving the objects on screen. Objects can be saved by clicking on the object and destroyed through using the destroy button. When you click to destroy or save an object, your response is recorded by the computer and the message 'Response Detected' will be displayed beside your score. PLEASE NOTE THAT YOUR SCORE IS DISPLAYED IN THE BOTTOM RIGHT HAND CORNER OF THE SCREEN. You will receive a prize of €5.00 in cash should you score more than 80 points in the course of the game. You must reach a score of 20 points in order to finish the first level of the game. When you are ready, click continue to begin.

The instructions presented before Levels 2 and 3 of both the Short Delay and Long Delay Games were as follows:

In this level, you must score as many points as possible in the time given. YOUR SCORE WILL NOT BE DISPLAYED until you have finished both levels two and three. However, although the computer is unable to display your score during the game, it will still display 'Response Detected' whenever you click to save or destroy an object. Remember that there is a prize of €5.00 in cash available to anyone who scores 80 points or more in the course of the game, so act as quickly and accurately as possible. When you are ready, click the continue button to begin the level.

As in Experiment 2.1, there were two responses available to participants on each trial, a save response and a destroy response. When either a save or destroy response was recorded, the character displayed was removed from the screen, the score displayed in the bottom right hand corner of the screen was adjusted accordingly and, unlike in Experiment 2.1, a 'Response Detected' message was displayed next to the score indicator. In addition, once any response was made, the 'DESTROY!' button immediately became grey or 'disabled' and remained so until the onset of the following trial.

Both the Short-Delay and Long-Delay Games contained simulated network delays, designed to functionally simulate the effects of network delays on on-line game playing. Simulated Network Delays were inserted on one quarter of all trials presented in Levels 2 and 3, and were evenly distributed across both stimulus sets. A delay trial was initially indistinguishable from a trial that did not contain delays. However, there were a number of differences that only became apparent after a response was made by the participant. More specifically, on a delay trial a character was presented in one of 24 possible locations on screen. That character then increased in size by 25% every 0.5s of the trial until the full 2s had elapsed, unless the participant made a response. If the participant made a response in a delay trial, a number of events occurred: 1) The character remained on-screen, but did not increase in size until the duration of the delay had passed (i.e., it froze); 2) the 'Response Detected' message was not displayed; and 3) the 'DESTROY!' button turned a grey colour, and became disabled for the duration of the delay before returning to its regular appearance and functionality for the remainder of the trial. In the Short Delay Game, the delay lasted for 0.5s, while in the Long Delay Game the delay lasted for 1s. During delays, participants were unable to make a response, either by clicking on the character or the 'DESTROY!' button. Both the returning of the 'DESTROY!' button to its regular appearance and the characters' resumed 'approach' was designed to signal to the participant that they could now produce a score if they made the correct response within the time remaining in that trial.

Both the Short Delay and Long Delay Games consisted of three levels; Level 1 was a Baseline Training Level in which participants learned how to interact with the interface; Level 2 required participants to demonstrate one node stimulus equivalence in order to achieve a high score; Level 3 required participants to demonstrate three-node stimulus equivalence in order to achieve a high score. It should be noted that the order of presentation of Levels 2 and 3 was counterbalanced across all conditions to eliminate any possible order effects.

In Level 1, which will hereafter be referred to as the Training Level, the characters presented in the game space were the 'A' stimuli from the Stimulus Equivalence Training and Testing phase. The functioning of each trial was very similar to Experiment 1. In this level, points were earned for

saving the A1 stimulus (a red circle) and destroying the A2 stimulus (a green square) within the 2s given for a response. It must be noted that no limit was placed on the number of trials to which a participant could be exposed during Level 1. Specifically, an incorrect response or a failure to respond within the given time limit resulted in a score of -1 for that trial. Thus, a potentially infinite number of trials was required in order for the participant to achieve a score of +20.

In Level 2, which will hereafter be referred to as the one-node equivalence level, the characters that were presented in the game space were the 'C' stimuli from the Stimulus Equivalence Training and Testing phase. Points were gained for responding in accordance with the established equivalence relations. For example, as points were gained in Level 1 for saving the A1 stimulus (a red circle), points were gained in Level 2 for saving the C1 stimulus (a yellow pentagon), which participated in a derived equivalence relation with the red circle. Similarly, just as points were gained in Level 1 for destroying the A2 stimulus (a green square), points were gained in Level 2 for destroying the C2 stimulus (a blue triangle). These novel equivalently-related stimuli were presented in Level 2 without warning, thereby requiring participants to engage in a degree of problem solving in order to achieve a high score without trial-by-trial feedback.

Level 3, which will hereafter be referred to as the three-node equivalence level, involved the presentation of the 'E' stimuli from the Stimulus Equivalence Training phase. Upon the completion of the 48 trials, Level 3 came to an end and the participants' score for both Levels 2 and 3 were presented on-screen.

It is important to remember that, as in Experiments 2.1 and 2.2, the Correct Responses measure was not confounded with the effects of the simulated truncated response window during delay trials. That is, Correct Responses made during a delay trial, while leading to a loss of points on the score displayed to participants, were actually recorded as Correct Responses for the purposes of data analysis (see 2.1.1.4 for more detail). In this way, it was possible for all participants in all conditions to achieve a Correct Response rate of 100%, irrespective of their score.

Upon the completion of each level of the game, a brief questionnaire, which consisted of an element of the Day Reconstruction Method (DRM; Kahneman, et al., 2004, see Appendix 1 for original instrument) designed to measure subjective responses to recent events, was presented on-screen. Two constructs, Positive Affect and Negative Affect, were derived from summing all of the items on the questionnaire relating to positive and negative experiences, respectively. These constructs represented the two subjective measures employed in the study. Twelve questions were presented and participants indicated their response by sliding a bar to the appropriate position on a six point Likert scale. Once all questions were answered the participant clicked a 'submit' button to

continue.

2.3.2 Results and Discussion

All participants passed the four stimulus equivalence training blocks within three exposures to each block. Eleven participants passed the Stimulus Equivalence Test on their first exposure, a further ten passed on their second exposure, one passed on their third exposure and only one participant failed to pass the Stimulus Equivalence Test within four exposures. Table 2.17 details performances on Stimulus Equivalence Training and Testing phases. Please note that where an ‘x’ appears under Testing, this means that the participant did not pass the equivalence test and required further training. The test performance marked with * indicates that the relevant participant passed on their third exposure to the training and testing procedure. The test performance marked with ** indicates that the relevant participant failed to pass on their fourth exposure to the training and testing procedure.

Table 2.17: Number of exposures required to pass each stage of Stimulus Equivalence Training and Testing

Participant	A-B	B-C	C-D	D-E	Test	A-B	B-C	C-D	D-E	Test
1	1	1	1	1	x	1	1	1	1	1
2	1	1	1	1	1					
3	1	1	3	1	1					
4	3	1	1	2	1					
5	2	1	1	1	x	1	1	1	1	x**
6	2	1	1	1	x	1	1	1	1	x*
7	1	1	1	2	x	1	2	1	2	1
8	3	1	1	1	x	1	1	2	1	1
9	1	1	1	1	1					
10	1	1	1	2	x	2	1	1	1	1
11	1	1	1	1	x	1	1	1	2	1
12	1	1	1	1	1					
13	1	1	2	1	1					
14	2	1	1	1	1					
15	1	1	1	1	x	1	1	1	1	1
16	1	1	1	1	1					
17	1	1	2	1	1					
18	1	1	1	3	x	1	1	1	1	1
19	1	1	1	1	x	1	1	1	1	1
20	1	1	1	1	1					
21	2	1	3	1	x	1	1	1	1	1
22	2	1	1	1	1					
23	2	1	1	1	x	1	1	1	1	1

Twenty two participants passed the Stimulus Equivalence Training and Testing phase and were

then assigned to play one of two games, one of which contained 0.5s simulated network delays and one of which contained 1s simulated network delays. All Participants were initially presented with a training level in which they learned how to interact with the game characters and novel graphic interface. As in Experiments 2.1 and 2.2 Correct Responses are not presented for this level.

2.3.2.1 Training level

Table 2.18: Mean scores for all measures employed on the Training Level of both the Short Delay (n=11) and Long Delay (n=11) Games.

	Short Delay	Long Delay
Positive Affect	8.272727	8.090909
Competence	4	3.909091
Negative Affect	2.090909	5.454545
Impatience	1.818182	1.818182

Table 2.18 presents mean scores for Positive Affect, Competence, Negative Affect and Impatience for the Training level. Please note that as there was no delay on any trials presented to either group in the Training Level, the two groups were actually presented with exactly the same task. As such, it is no surprise to see that scores are very similar across the two groups on most of the measures presented. Interestingly, however, the participants who were later presented with the Long Delay Game rated the training level as higher in Negative Affect than the group who were later presented with the Short Delay Game.

2.3.2.2 Short Delay Game

Correct Responses were calculated as the total number of correct first responses to trials in a given level. It is important to remember that, as in Experiments 2.1 and 2.2, there is a distinction between the score displayed to participants during game play and the total number of Correct Responses made (see section 2.1.1.4 for more detail). Only the latter is used in the following analysis. Table 2.19 presents the Correct Responses made by each participant across the one-node and three-node Levels in the short Delay Game.

It is evident by the number of Correct Responses made that most participants demonstrated robust relational responding in both the one-node and three-node stimulus equivalence levels. Eight of the eleven participants who played the short Delay Game produced 90% correct responding in the one-node level, while six participants produced 90% correct responding in the three-node level.

Table 2.19: Correct Responses made by each participant across the one-node and three-node Levels in the short Delay Game

Short Delay											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
1 Node	48	3	44	0	48	47	45	47	47	47	4
3 Node	0	1	46	1	47	1	48	48	48	47	4

2.3.2.3 Long Delay Game

Table 2.20: Correct Responses made by each participant across the one-node and three-node levels in the long Delay Game

Long Delay											
	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
1 Node	47	48	48	47	48	48	48	48	48	48	48
3 Node	47	0	0	48	48	47	1	47	48	47	48

As displayed in Table 2.20, most participants who played the Long Delay Game also appeared to demonstrate robust relational responding in both the one-node and three-node stimulus equivalence levels. All eleven participants who played the Long Delay Game produced 90% correct responding on the one-node level, while eight participants produced 90% correct responding in the three-node level. Interestingly, more participants reached the 90% criterion while playing the long Delay Game than the Short Delay Game. However, due to low participant numbers it is difficult to decipher whether this is due to the delay itself, or individual differences between the two participant groups.

An examination of the raw Correct Responses reveals that participants appeared to either reach a 90% correct responding criterion or else produce very few Correct Responses. Correct Responses in the range of 0-4/48 suggest that participants responded consistently incorrectly. In effect, it appears that all participants made consistent stimulus discriminations in almost all trials across both levels and both games, but that these discriminations were not always consistent with the previously established equivalence relations.

Table 2.21 presents mean Correct Responses, mean Positive Affect and Negative Affect, mean Impatience scores and mean competence scores across all three levels of both games. Positive Affect, Competence, Negative Affect and Impatience are constructs measured by an element of the Day Reconstruction Method (DRM; Kahneman et al., 2004) questionnaire. A total of twelve questions were posed to participants on the computer screen after the completion of each level of the game. Answers were made using a seven-point Likert scale. Three of these items were summed in order to produce a Positive Affect rating, while six of the remaining items were summed to generate a

Negative Affect rating. Competence and Impatience were individual items on the questionnaire.

Table 2.21: Means for all measures employed across the Training, One-Node and Three-Node Levels, across the Short Delay (n=11) and Long Delay (n=11) Games.

	Training	One-Node	Three-Node
Correct Responses			
Short Delay	N/A	34.5	36.5
Long Delay	N/A	47.9	34.6
Positive Affect			
Short Delay	8.272727	6	5.727273
Long Delay	8.090909	6.909091	7.818182
Long Delay			
Short Delay	4	2.181818	3.090909
Long Delay	3.909091	3	2.909091
Negative Affect			
Short Delay	2.090909	4.636364	4.636364
Long Delay	5.454545	6.363636	6
Impatience			
Short Delay	1.818182	2.181818	2.090909
Long Delay	1.818182	2.363636	1.727273

Table 2.21 illustrates the mean scores recorded for all measures employed in the current experiment. Relational Complexity does not appear to have had a consistent effect on Correct Responses. Correct Responses increase from the One-Node Level to Three-Node Level in the short delay game, but decrease across the same levels in the Long Delay game. The effect of Relational Complexity on Positive Affect ratings is also not consistent across all conditions. In both games there is a decrease in mean positive affect ratings from the Training Level to the One-Node Level. However, mean ratings then decrease further in the Short Delay Game, while increasing in the Long Delay Game. A similar pattern is evident in mean Negative Affect scores. Specifically, there is an increase in Negative Affect scores from the Training Level to the One-Node Level, suggesting participants enjoyed One-Node Level less than the Training Level. However, the pattern across One-Node and Three-Node levels is less clear. Similarly ambiguous patterns are observed for ratings of competence and impatience. Thus, it appears that while participants prefer a game which does not require derived relational responding, the level of complexity involved in the Three-Node Level is not less enjoyable to participants than that involved in the One-Node Level.

A bivariate correlational analysis was conducted on the variables of game level (i.e., levels 1, 2 or 3), and ratings of Positive Affect, in order to better understand the relationship of enjoyment and game difficulty. Specifically, the relationship between Positive Affect and game level (i.e., the training, One-Node and Three-Node levels) was investigated using a Pearson product-moment

correlation coefficient. There was no significant correlation between the two variables [$r=-0.134$, $n=66$, $p=0.141$]. Thus, the relationship between positive affect and complexity, at least in the current experiment, seems unclear.

Network Latency does not appear to have had any consistent effect on Correct Responses. Correct Responses are higher in the Long Delay Game than the Short Delay Game at the One-Node Level, but lower in the Long Delay Game at the Three-Node Level. Mean Positive Affect ratings are higher in both the One Node and Three Node levels of the Long Delay Game, while there is only a small difference in mean ratings in the Training Level, suggesting that the game containing more latency was more enjoyable for participants. However, this pattern was not reflected in mean Negative Affect ratings. Negative Affect ratings were higher for the long delay game in all three levels of the long delay game. Thus, participants appeared to rate the Long Delay Game as both more positive and more negative than the Long Delay Game. There was no consistent pattern for ratings of Competence or Impatience.

One of the principal aims of the study was to examine whether there was an interaction between the variables of Relational Complexity and Network Latency on both Correct Responses and subjective ratings. Comparing mean ratings of Positive and Negative Affect across the One Node and Three Node levels, it appears as if there may be an interaction between the two variables of interest. Specifically, ratings of Positive and Negative Affect are relatively consistent in the Short Delay Game. Positive Affect decreases consistently across levels, while negative affect increases across levels. However, while the same pattern is observed from Training Level to One-Node Level of the long delay game, the same pattern is not observed across the One Node and Three Node levels. Mean positive affect ratings increase from the One-Node Level to Three-Node Level and Negative Affect ratings decrease for the same levels. It appears as if the extra physical challenge presented by the long delays, in conjunction with the higher relational challenge of Three-Node Level, may have lead to a more enjoyable experience for participants. This conclusion is supported by ratings of impatience, which drop from the One-Node Level to the Three-Node Level in both game types. These trends were analysed using an inferential statistical test in order to determine whether the patterns observed were significant.

Table 2.22 displays the results of five mixed between-within subject's analysis of variance, conducted to explore the impact of Relational Complexity and Level of Delay on participants' Correct Responses and ratings of Positive and Negative Affect, as well as ratings of competence and impatience. Those p values marked with an asterisk represent significant effects. Delay Level had no significant effect on any measure employed in the current study. It would appear that both the 0.5s

Table 2.22: Results from a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=22)

	Wilks' Lambda	F Value	p Value	Eta Squared
Correct Responses				
Complexity	0.774	5.834	0.025*	0.226
Delay	N/A	2.403	0.137	0.107
Interaction	0.983	0.343	0.565	0.017
Positive Affect				
Complexity	0.637	5.405	0.014*	0.363
Delay	N/A	0.298	0.591	0.015
Interaction	0.902	1.027	0.377	0.098
Competence				
Complexity	0.524	8.635	0.002*	0.476
Delay	N/A	0.063	0.805	0.003
Interaction	0.903	1.026	0.378	0.097
Negative Affect				
Complexity	0.830	1.939	0.171	0.170
Delay	N/A	1.387	0.253	0.065
Interaction	0.948	0.523	0.601	0.052
Impatience				
Complexity	0.922	0.806	0.461	0.078
Delay	N/A	0.010	0.920	0.001
Interaction	0.959	0.402	0.675	0.041

and the 1s delay periods produced similar levels of positive and negative affect, as well as competence and impatience. This would suggest that, at least using the current preparation, extended delays of varying lengths are not distinguishable from each other in terms of their influence on positive and negative affect. Additionally, there were no significant interactions between the variables of relational complexity and delay level in the current experiment.

Relational Complexity had a significant effect on Correct Responses and ratings of positive affect and competence. However, relational complexity did not have a significant effect on ratings of negative affect or impatience. The finding that participants produced a lower number of Correct Responses in the later (more complex) levels than in the earlier (less complex) levels, suggests that participants had greater difficulty in making the appropriate response in the more complex levels. This provides support for the selection of nodal distance as an appropriate variable in the manipulation of relational complexity, despite its ambiguous effect on reported enjoyment levels

Perhaps the most unexpected finding of Experiment 2.3 was that Relational Complexity had a significant influence on Positive Affect ratings in the opposite direction to that predicted. The more relationally complex games were rated as significantly less positive by participants. Given these

results we must consider the possibility that complexity may be a non-significant factor in determining game enjoyment. Alternatively, complexity may be detrimental to the game playing experience. However, the scientific process demands that we reach this conclusion only after a more detailed effort to manipulate our key variables and have achieved increased levels of control over them.

More specifically, experimental control may have been compromised in Experiment 2.3 by the administration of an equivalence test prior to the game stage. Such a test requires participants to derive both the one-node and three-node relations in advance of the game. Therefore, the procedure employed in Experiment 2.3 may have negated the need for participants to derive the appropriate relations during game play itself. In effect, if subjects commenced game play with the equivalence relations already fluently established, the games would consist merely of a tedious memory exercise involving two equivalence relations and ten stimuli, rather than a genuine dynamic derived relational responding exercise. Given this possible interpretation, we might have expected to find that subjects did not prefer the Three-Node Level over the One-Node Level.

In addition to the foregoing, the fact that the equivalence test only probed for three-node A-E relations, and not A-C relations, meant that participants had a more established history of responding to the derived E-A relations than to C-A relations. Thus, although the Three-Node equivalence level should have been more difficult than the One-Node Level, the greater amount of practice participants had with the three-node relations may have minimized differences across the game levels. One way to test the foregoing possibility is to repeat Experiment 2.3, omitting the presentation of an equivalence test before participants are exposed to the game. This is exactly the procedure employed in Experiment 2.4.

2.4 Experiment 2.4

In Experiment 2.3, the majority of participants demonstrated both one-node and three-node equivalence relations in the context of a computer game in which there are significant response time demands. Relational complexity was found to significantly affect total Correct Responses in the game. Specifically, participants attained lower scores in the Three-Node Level than in the One-node Level. Relational complexity also significantly affected positive affect ratings in Experiment 2.3, insofar as participants seemed to prefer less complex over more complex games. More specifically, participants rated the more complex levels as significantly lower in positive affect than the less complex levels. Participants also rated themselves as significantly less competent at the more

complex levels. Relational Complexity had no effect on negative affect or impatience scores. Network Latency did not have a significant effect on any measures employed in Experiment 2.3.

These findings may appear to suggest that the 0.5s and 1s Delay Games are indistinguishable from each other in their effects, and that relationally more complex games are less enjoyable than less complex games. However, in order to confirm these findings a further experiment must be conducted, in which even greater control is exerted over the variable of relational complexity. In particular, experimental control may have been compromised in Experiment 2.3 by the administration of an equivalence test prior to the game stage. In order to pass such a test, participants must demonstrate fluent responding over a large number of trials, in essence demonstrating that they have already solved the problem posed in the game (i.e. to derive equivalence relations) Thus, even though Experiment 2.3 involved a higher level of complexity than the experiments reported thus far in the thesis, it may have failed to generate the very high levels of relational and problem-solving complexity required to observe the ameliorative or detrimental effect of complexity on game performance and enjoyment. For this reason, Experiment 2.4 involved repeating Experiment 2.3, with the omission of an equivalence test. Participants first received equivalence training, identical to that presented in Experiment 2.3. Once equivalence training was completed, participants proceeded directly to the game stage, which was identical to that employed in Experiment 2.3. As in Experiment 2.3, a questionnaire was presented on-screen after the completion of each of the three levels in order to measure participants' subjective evaluations of those levels.

2.4.1 Method

2.4.1.1 Participants

Twenty two participants were recruited, all of whom were undergraduate students. Eleven of these were male, while 11 were female. Participants were offered a payment of €5 upon completion of the game.

2.4.1.2 Materials

The materials used were the same as those used in Experiment 2.3.

2.4.1.3 Design

The Design of experiment 2.4 was identical to that of Experiment 2.3.

2.4.1.4 Procedure

As in Experiment 2.3, participants were first exposed to the matching-to-sample stimulus discrimination procedure, in which two three node equivalence classes (A1-B1-C1-D1-E1 and A2-B2-C2-D2-E2) were established. Training was conducted sequentially, in blocks of 20 trials where a criterion of 19 Correct Responses was employed. However, unlike in Experiment 2.3, once participants had successfully completed all four training phases (A-B, B-C, C-D, and D-E) they did not then receive a stimulus equivalence test. Instead, participants advanced immediately to the game stage, which was identical to that employed in Experiment 2.3.

2.4.2 Results and Discussion

Table 2.23 details performances on Stimulus Equivalence Training. All participants passed the four stimulus equivalence training blocks within three exposures to each block.

Table 2.23: Number of exposures required to pass each stage of Stimulus Equivalence Training

Participant	A-B	B-C	C-D	D-E
1	2	1	1	1
2	3	2	1	1
3	1	1	1	1
4	1	1	1	2
5	1	2	1	1
6	2	1	1	1
7	1	1	2	1
8	3	1	2	1
9	1	2	1	1
10	2	2	2	1
11	7	2	3	1
12	1	1	1	2
13	1	1	1	1
14	1	2	1	1
15	1	2	1	1
16	1	1	1	1
17	1	2	1	1
18	1	1	1	1
38	1	1	1	2
19	1	1	1	1
20	2	1	1	1
21	3	2	1	1
22	1	1	1	1

All participants passed Stimulus Equivalence Training and Testing and were assigned to play one of two games, one of which contained 0.5s simulated network delays and one of which contained

1s simulated network delays. All participants were initially presented with a training level in which they learned how to interact with the game characters and novel graphic interface. As in Experiments 2.1-2.3 Correct Responses are not presented for this level.

2.4.2.1 Training level

Table 2.24: Mean scores for all measures employed on the Training Level of both the Short Delay (n=11) and Long Delay (n=11) Games.

	Short Delay	Long Delay
Positive Affect	8	7.818182
Competence	3.363636	4.272727
Negative Affect	5.454545	4.818182
Impatience	1.636364	1.454545

Table 2.24 presents mean scores for Positive Affect, Competence, Negative Affect and Impatience for the Training level. Please note that as there was no delay on any trials presented to either group in the training level, the two groups were actually presented with exactly the same task. As such, it is no surprise to see that scores are very similar across the two groups on most of the measures presented.

2.4.2.2 Short Delay Game

Table 2.25: Total Correct Responses made by each participant across the One-Node and Three-Node Levels in the Short Delay Game

		Short Delay										
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
1 Node		48	0	2	3	18	20	48	47	44	48	5
3 Node		48	9	0	1	6	21	1	47	35	47	46

Table 2.25 presents the number of Correct Responses made by each participant across the one-node and three-node levels in the short Delay Game. Five of the eleven participants who played the Short Delay Game produced 90% correct responding in the one-node level, while five participants also produced 90% correct responding in the three-node level. Thus, it does not appear that participants were more successful in responding to trials in either the one-node or three-node levels.

2.4.2.3 Long Delay Game

Table 2.26 presents the number of Correct Responses made by each participant across the one-node and three-node Levels in the Long Delay Game. Three participants who played the Long

Table 2.26: Total Correct Responses made by each participant across the One-node and Three-node Levels in the Long Delay Game

Long Delay											
	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
1 Node	2	23	40	48	47	48	0	0	13	4	26
3 Node	47	21	3	43	0	48	48	5	41	16	19

Delay Game produced 90% correct responding in the one-node level, while four participants produced 90% correct responding in the three-node level. Interestingly, fewer participants reached the 90% criterion while playing the long Delay Game than the short Delay Game. However, due to low participant numbers it is difficult to decipher whether this is due to the delay itself, or individual differences between the two participant groups.

An examination of the raw Correct Responses reveals that a large number of the participants who failed to demonstrate 90% correct responding produced very few Correct Responses. Typically three or four of the eleven participants who played each level recorded between zero and four Correct Responses. Correct Responses in the range of 0-4/48 suggest that participants responded consistently incorrectly. In effect, it appears that all participants made consistent stimulus discriminations in almost all trials across both levels and both games, but that these discriminations were not always consistent with the previously established equivalence relations. Importantly, these findings are similar to those of Experiment 2.3.

Table 2.27: Means for all measures employed across the Training, One-Node and Three-Node Levels, across the Short Delay (n=11) and Long Delay (n=11) Games

	Training	One-Node	Three-Node
Correct Responses			
0.5s Delay	N/A	25.7	23.7
1s Delay	N/A	22.8	26.5
Positive Affect			
0.5s Delay	8	7.545455	5.272727
1s Delay	7.818182	6.727273	5.090909
Competence			
0.5s Delay	3.363636	2.545455	2.363636
1s Delay	4.272727	2.818182	2.545455
Negative Affect			
0.5s Delay	5.454545	5.727273	8.090909
1s Delay	4.818182	5.545455	6.454545
Impatience			
0.5s Delay	1.636364	2.636364	2.727273
1s Delay	1.454545	2	2.818182

Table 2.27 illustrates the mean scores recorded for all measures employed in Experiment 4. Relational Complexity does not appear to have had a consistent effect on Correct Responses. Correct Responses decrease from the One-Node Level to the Three-Node Level in the Short Delay Game and increase across the same levels in the Long Delay Game. Participants appeared to rate the more complex levels as lower for Positive Affect. In both game types there is a steady decrease in Positive Affect ratings across all three levels, indicating that participants prefer the less complex levels of the game. Furthermore, the inverse pattern is observed for ratings of Negative Affect. In both game types there is a steady increase in negative affect ratings across all three levels, again indicating that participants prefer the less complex levels of the game. In addition, ratings of Competence decrease consistently across levels while ratings of Impatience increase. Thus, it appears that participants consistently rated the levels requiring a higher level of derived responding as less enjoyable than less relationally complex levels.

A bivariate correlation was conducted on the variables of game level (i.e., levels 1, 2 or 3), and ratings of positive affect, in order to better understand the relationship of enjoyment and game difficulty. Specifically, the relationship between Positive Affect and game level (i.e., the Training, One-Node and Three-Node levels) was investigated using a Pearson product-moment correlation coefficient. There was a small negative correlation between the two variables [$r = -0.226$, $n=66$, $p=0.034$], with lower levels of positive affect related with later more complex levels of the game.

Network Latency does not appear to have had any consistent effect on Correct Responses. Correct Responses are higher in the Short Delay Game than the Long Delay Game at the One-Node Level, but lower in the Short Delay Game at the Three-Node Level. Mean Positive Affect ratings are higher for the Short Delay Game across all three levels, suggesting that the game containing less latency was more enjoyable for participants. However, this pattern was not reflected in mean Negative Affect ratings. Mean Negative Affect ratings are higher for the short delay game across all three levels. Thus, there appears to be no consistent preference by participants for one level of delay over another within the context of a game based on stimulus-equivalence problem solving.

One of the principal aims of the study was to examine whether there was an interaction between the variables of Relational Complexity and Network Latency on both Correct Responses and subjective ratings. From inspecting the means, there does not appear to be any interactions between these variables for Experiment 4.

Table 2.28 presents results from a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Total Correct Responses, Positive Affect, Competence, Negative Affect and Impatience. Those marked with an asterisk represent significant

Table 2.28: Results from a mixed between-within subjects ANOVA, testing for the effects of delay level and relational complexity on Total Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=22)

	Wilks' Lambda	F Value	p Value	Eta Squared
Correct Responses				
Complexity	0.999	0.023	0.882	0.001
Delay	N/A	0.000	0.990	0.000
Interaction	0.987	0.270	0.609	0.013
Positive Affect				
Complexity	0.674	4.598	0.024*	0.326
Delay	N/A	0.045	0.834	0.002
Interaction	0.993	0.070	0.933	0.007
Competence				
Complexity	0.608	6.127	0.009*	0.392
Delay	N/A	0.363	0.553	0.018
Interaction	0.954	0.461	0.637	0.046
Negative Affect				
Complexity	0.883	1.253	0.308	0.117
Delay	N/A	0.178	0.678	0.009
Interaction	0.972	0.273	0.764	0.028
Impatience				
Complexity	0.751	3.153	0.066	0.249
Delay	N/A	0.116	0.737	0.006
Interaction	0.971	0.283	0.757	0.029

effects. Relational Complexity had a significant effect on participants' ratings of positive effect. Specifically, participants found the later (more complex) levels of the game significantly less positive than the earlier (less complex) levels. Moreover, participants rated themselves as less competent with the more complex levels than the less complex levels. However, Relational Complexity did not have a significant effect on either Negative Affect, Impatience or Correct Responses. Delay Level had no significant effect on any measure employed in the experiment. In addition, there were no significant interaction effects between the variables of Relational Complexity and Delay Level.

It was proposed in the introduction to Experiment 2.4 that participants in Experiment 2.3 may not have been engaging in relational responding, per se, while playing the games, due to the administration of equivalence test prior to the game stage. This may explain why participants rated the less complex games as more enjoyable than the more complex games in Experiment 2.3. However, similar results to those found in Experiment 2.3 were found in Experiment 2.4, despite the omission of an equivalence test from the procedure. Thus, it would appear once again that participants prefer less complex over more complex games.

Interestingly, the number of participants achieving high scores is low in Experiment 2.4

compared to Experiment 2.3. Perhaps the administration of an equivalence test after the game stage in Experiment 2.4 would have clarified whether the equivalence classes were as well established for participants in Experiment 2.4 as those in Experiment 2.3. It is possible that participants who did not achieve high scores in Experiment 2.4 would also have failed to produce the derived relations during a test phase. That is, training alone may not be enough to generate competent playing for most participants. In effect, it appears that information required to play a game needs not only to be administered, but also to be consolidated before a game can be played competently by most participants.

An alternative explanation for poor game performances may be found in the training and testing procedure of Experiment 2.3. Specifically, participants who failed an equivalence testing block in Experiment 2.3 were re-exposed to the training procedure before being presented with a further testing block. Eleven of the twenty-two participants in Experiment 2.3 failed to reach criterion on their first exposure to an equivalence testing block, but passed on further exposures. Thus, eleven participants in Experiment 2.3 received more training than all participants in Experiment 2.4. The longer history of equivalence training received by participants in Experiment 2.3 may explain their better performance on the subsequent games.

Delay Level had no significant influence on Total Correct Responses, Positive affect or Negative Affect in Experiment 2.4. This replicates the findings of Experiment 2.3, despite the procedural variation employed. Thus, we may conclude that, at least with the current procedure, extended delays of varying lengths are not distinguishable from each other in terms of their influence on Positive and Negative Affect. However, it must be noted that, because there was no control condition (i.e., no delay) condition employed in either Experiment 2.3 or 2.4, the current results may be qualified. More specifically, these findings may not suggest that network latency has no effect on participants' performances or enjoyment of games more generally. This is because we do not know what effect, if any, network latency had on performances and subjective ratings in the current study. It is entirely possible that, under a 'no delay' condition, participants would have preferred the more relationally complex levels.

2.5 General Discussion

The current experiments represent the first step in a systematic behavioural investigation of on-line computer gaming. They demonstrate processes of generalization and derived relational responding in the context of a simple computer game format and thus show how important features

of computer game playing may be understood and analysed in behavioural terms. In addition, a definition of game complexity in terms of nodal distance was advanced and tested in Experiments 2.3 and 2.4. The four experiments reported here also examined the effect of network delay on players' performance as well as their subjective assessment of game quality.

All participants in Experiments 2.1 and 2.2 who completed the Training Level scored at above chance in the Generalisation phase. This finding demonstrates stimulus generalisation in the context of game playing and it is offered here as a suitable analogue of some low-level skills required by rudimentary games. In addition, all but two participants in Experiments 2.1 and 2.2 who completed the training phase scored above chance levels in the equivalence phase. In order to gain a high score in the latter phase, participants were required to demonstrate a transfer of the appropriate response function through derived equivalence relations. The finding that a large number of participants were able to consistently respond in accordance with equivalence relations provides support for the view that game players can show derived relational responding in the context of computer games.

The current study also found that simulated network latency, as variously defined across Experiments 2.1 and 2.2, had a significant effect on level of correct responding as well as on subjective measures of difficulty and frustration. Participants found the Delay Game significantly more difficult and frustrating than the Non-Delay Game, and achieved significantly lower scores when playing the former than the latter. Moreover, it is important to appreciate that these effects were replicated across two separate experiments, thus suggesting that these findings are quite reliable and easily approached and analysed using the current derived relations procedures.

Despite the foregoing, simulated network latency was not found to influence participants' enjoyment ratings of the games. This unexpected outcome could be interpreted in two ways. Firstly, it may be suggested that enjoyment of a game is uncorrelated with success at that game, or the difficulty and frustration experienced. However, this interpretation would appear to contradict a number of research findings (e.g., Malone, 1982; Vorderer, Hartmann, & Klimmt, 2003), which suggest that the level of difficulty of a game and enjoyment experienced with that game are closely related.

Secondly, it could be argued that reliable and valid data is always difficult to obtain using non-standardised subjective rating scales such as those employed in the current study. It must be remembered, however, that the current research is highly novel and exploratory and is intended to pave the way for the development of reliable subjective measures that may be validated against objective empirical information regarding the experience of on-line gaming. In effect, only by understanding the experience of game playing at a quantitative level can we hope to develop the

kinds of subjective measures that will be of interest to both psychologists and their research counterparts in the gaming industry. This issue will be pursued in further experiments presented in the current thesis.

Perhaps the most interesting finding of Experiments 2.1 and 2.2 was that neither experiment found a significant effect for relational complexity on game performance or on any of the subjective measures. In effect, as complexity increased across levels of the game, participants' performance and experience of the game were unaffected. This finding suggests that complexity, at least as conceived in Experiments 2.1 and 2.2, does not alter the experience of gaming by increasing levels of enjoyment, difficulty or frustration.

Alternatively, the finding that relational complexity had no effect on game performance or on any of the subjective measures may suggest that complexity was not successfully manipulated in Experiments 2.1 and 2.2. It could be argued that the levels of relational responding examined in these experiments were relatively low compared to those required in a large number of modern games. In effect, it might not be that participants enjoy less relationally complex games across the board, but appear to do so only when the levels under examination are low and relatively non-stimulating.

Experiments 2.3 and 2.4 were conducted in order to address three issues which arose in the analysis of Experiments 2.1 and 2.2. Firstly, Experiments 2.3 and 2.4 aimed to extend the analysis of relational complexity conducted in Experiments 2.1 and 2.2 by increasing the complexity of the relational activities required to play the games. This was carried out through manipulating the size of the stimulus classes in both experiments. Additionally, the stimulus equivalence testing phase was removed from the procedure of Experiment 2.4 in an attempt to clarify whether the administering of an equivalence test before game play was a confounding factor on the impact of relational complexity on game enjoyment.

Secondly, Experiments 2.3 and 2.4 were designed to extend the analysis of network latency carried out in Experiments 2.1 and 2.2, by extending the length of delays in the games. This strategy was adopted due to findings from the field of engineering (i.e., Delaney, Meenaghan, Ward, & McLoone, 2004; Vaghi, Greenhalgh, & Benford, 1999), which suggest that incrementally increasing the level of delay in a game leads to a progressively less 'playable' game.

Thirdly, it was unclear in Experiments 2.1 and 2.2 whether there genuinely was no link between network latency and enjoyment, or whether the results were occluded to some extent by the non-standardised rating scales employed. Thus, a rating scale was employed in Experiments 2.3 and 2.4 that was developed in order to measure subjective experience of recent events. Due to the fact

that extensive research had been carried out on this instrument, it should represent a more reliable subjective measure of enjoyment than previously employed.

Similarly to Experiments 2.1 and 2.2, participants in both Experiments 2.3 and 2.4 demonstrated robust derived relational responding in a game playing environment. Indeed, participants in Experiment 2.3 had relatively little difficulty in reaching a 90% correct response criterion on the three-node equivalence level. This behaviour, therefore, may serve as a model of the types of complex behaviours observed in some online computer games.

Relational Complexity significantly affected scores on Experiment 2.3. Specifically, participants attained lower total Correct Responses in the later levels of the games than in the earlier levels. This finding confirms that the complexity variable was manipulated correctly, insofar as higher levels of complexity should lead to lower scores. Thus, these findings confirm that through manipulating the level of derived relational responding required in a game, the level of difficulty or challenge experienced by the players can be increased. As such, this finding supports the derived relations approach to understanding users' experience of online games and confirms that difficulty of a game can be manipulated experimentally in a quantifiable and functionally understood way.

It was argued that experimental control may have been compromised in Experiment 2.3 by the administration of an equivalence test prior to the game stage. Such a test requires participants to derive both the one-node and three-node relations in advance of the game. Therefore, the procedure employed in Experiment 2.3 may have negated the need for participants to derive the appropriate relations during the game play itself. As such, a stimulus equivalence test was not administered before game play in Experiment 2.4.

In Experiment 2.4, no significant difference was found between scores on the earlier levels of the game and those on the later levels. This finding may be explained due to the relatively high demand on participants in Experiment 2.4 compared to Experiment 2.3. This high demand was generated by the absence of a stimulus equivalence testing phase in the procedure of Experiment 2.4. In effect, participants in Experiment 2.4 were required to engage in a complex problem-solving task under demanding response time constraints. In effect, this may have led to a floor effect on scores in both levels 2 and 3 in Experiment 2.4.

Relational Complexity significantly affected Positive Affect ratings in both Experiments 2.3 and 2.4. Closer inspection of the mean ratings, however, reveals that in Experiment 2.3, while there was a general trend of lower positive affect ratings for later levels, this trend was not reflected in the three-node level of the Long Delay Game. In the Long Delay Game Level 3 was rated as higher in positive affect than earlier levels. In addition, and contrary to expectations, relational complexity

did not have a significant effect on negative affect scores in Experiment 2.3. Similarly to those of Experiment 2.3, results of Experiment 2.4 suggest that participants appeared to not necessarily prefer more complex over simpler levels of the game. In Experiment 2.4, the later levels of the games were rated as significantly lower in positive affect than the earlier levels.

Experiments 2.3 and 2.4 were conducted in order to extend the analysis of network latency conducted in Experiments 2.1 and 2.2. Specifically, participants were presented with a game containing either short 0.5s delays or long 1s delays. No significant difference was found between these groups, on any measure employed in the study. This finding would suggest that there is no functional distinction between the short, 0.5s and long, 1s delays modelled in the current study. However, results of Experiments 2.3 and 2.4 must be viewed with caution, as there was no control or 'no delay' condition. These results, therefore, do not suggest that network latency has no effect on participants' performances or enjoyment of games more generally. In fact, given that we have no baseline scores against which to compare ratings, we do not know what effect, if any, network latency had on performances and subjective ratings in the current study. It is entirely possible that, under a 'no delay' condition, participants would have preferred the more relationally complex levels.

In Experiments 2.1 and 2.2, ratings of difficulty and frustration were significantly higher for participants who played a game containing simulated network delays than for those who played a game containing no delays. In addition, participants who played the game containing delays achieved significantly lower scores on that game. Thus, the presence of simulated network delays appears highly destructive towards the game playing experience. Taken together, the results of these two studies may suggest that while the presence of delay in a game is destructive to the game playing experience, increasing the length of that delay beyond 0.5s does not have any further negative effects.

The finding that a large number of participants were able to consistently respond in accordance with the previously learned equivalence relations provides support for the view that game players can show derived relational responding in the context of computer games. Indeed, this demonstration may serve as a suitable analogue of some types of complex skills required in many commercially available games. Furthermore, this analysis could provide the basis for creating educational games which combine the need to respond to the physical features of stimuli with the need to respond to the relationships between them. Importantly, using a derived relations approach would allow for the level of complexity of such games to be tightly defined.

It is important to note that there does not appear to be a link between the scores that participants achieved while playing the games in both Experiments 2.3 and 2.4 and their subjective ratings of those games for Positive and Negative Affect. In Experiment 2.3, there were significantly

lower scores observed in the more complex levels of the game and this was accompanied by lower positive affect ratings. However, Negative Affect ratings remained stable across levels. In Experiment 2.4, scores were not significantly different across levels, but there were significantly lower levels of Positive Affect reported for the more complex levels. In addition, Negative Affect ratings were higher in the more complex levels, but this difference was not significant. Thus, it appears that the scores participants achieve in a game do not necessarily determine their enjoyment of that game. At least in the current preparation, it appears that the two dependent variables which represent success at a game and enjoyment of that game are independent. This, in itself, is an important empirical finding for the computer games industry and the field of psychology more generally.

The finding that participants prefer to engage in less challenging games may of considerable interest to the computer games industry. Specifically, these findings challenge the conventional wisdom in the engineering literature that suggests a link between higher levels of challenge and enjoyment in games. It must be pointed out that challenge, or complexity, were defined only in relational terms in the current study. However, this in itself represents an important contribution. More specifically, while people on the whole may prefer more challenging to less challenging games, they do not necessarily want challenge to be presented in terms of relational complexity. It is possible that other forms of complexity, such as time constraints, precise motor skills, the number of stimuli involved, and so on, may make for an enjoyable game before relational complexity.

Of course, it could be argued that even the levels of relational responding examined in the current study were relatively low compared to those required in a large number of modern games. In effect, it might not be that participants enjoy less relationally complex games across the board, but appear to do so only when the levels under examination are low and relatively non-stimulating. The idea that the games presented in the current chapter were not sufficiently complex is borne out by the high number of Correct Responses observed in all games at all levels. Specifically, in a typical equivalence test format most researchers find a very low correct responding on the first exposure to the test, even in a context in which there is an unlimited time frame for responding. Thus, it would appear that the current game-based presentation format actually facilitated high levels of accurate responding which could be called exceptional by those experienced in the field of equivalence training and testing. It remains to be seen what enjoyment participants would derive from a game in which clearer evidence of a struggle to produce the correct answer under time constraints was evident. Future research should involve more complex relational activities such as responding in terms of relations of 'same,' 'opposite,' 'more than,' 'less than,' and so on, in order to more closely examine the relationship between complexity and enjoyment.

Despite the complexity of the issues arising from the current findings, game developers may do well to take heed of the possibility that relational complexity may not be a significant psychological variable in the understanding of online games, at least when manipulated at low levels. Of course, it is still expected that some degree of challenge is necessary for a game to be experienced as enjoyable. Indeed, this assertion is in line with research which suggests that participants most enjoy games with an appropriate but measured level of ‘challenge’ (e.g. Vorderer et al., 2003). However, no technical or psychological definition for the term ‘challenge’ is available in the literature and it is unclear whether or not the most enjoyable degrees of ‘challenge’ are relational or merely spatial and temporal. For instance, games designers may respond to the need for challenge by manipulating a wide variety of differing and possibly unrelated variables such as perceptual demands, the quantity of stimuli (e.g., game characters), the ratio of reflex responses to conscious responses required, the speed of stimulus presentation and the duration of response opportunity windows. Only a conceptually sophisticated theoretical framework and a sufficiently technical research methodology such as that provided by RFT will allow us to accurately distinguish the various dimensions of complexity that may be conceived. In so doing, researchers will be in a position to compare the relative effects of various dimensions of complexity on the gaming experience.

The four experiments presented in the current chapter were designed to analyse the functional relationships between the variables of relational complexity and network latency, and to assess the effects of these variables on a number of measures. The analysis has revealed that; 1) the effects of complexity may be observed in only a narrow range of game types that are neither very high nor low in complexity, and 2) that network delay appears to interact in a very complex way with complexity. Given the unclear nature of data gathered in the experiments conducted thus far, it is possible that there may be no consistent linear or easily ascertained relationship between the variables of relational complexity and network delay. Indeed, while it may be possible to disentangle the effects of these variables through refining the previous experiments, such an exercise may prove time consuming and fruitless. So, it may not be worthwhile to simply continue to pursue the latency / complexity interaction relationship until at least each of these variables has been examined more carefully independently of each other. Such a strategy will allow for a more measured and focused exploration of these variables as independent factors in game performance and enjoyment.

It has been argued that the level of relational responding presented in the current chapter may not have been sufficient to be enjoyable to participants. It is possible that in a game that requires more complex relational responding, the higher levels of complexity would be more enjoyable to participants. However, the method of increasing complexity in Experiment 2.4 (i.e., removing the

stimulus equivalence test from the training and testing procedure) lead to a game that was unplayable to most participants. An alternative method of presenting increased complexity may prove more fruitful for the investigation of the effect of complexity on enjoyment in games. One possible method of manipulating relational complexity, which would allow for the functional definition of a number of different levels of complexity, is to require participants to engage in complex relational responding. Future studies may benefit from such an approach.

Future studies should also aim to model network latency in a more ecologically valid format than achieved here. In Experiments 2.1 - 2.4, network latency was modeled as a fixed interval of either 0.5 seconds or 1 second. It may be argued that, given that participants could predict the length of each delay suffered, the delays could have been perceived as a particular challenge of the game, rather than a nuisance or problem with the game. In practice, network latency is rarely, if ever, predictable, and typically oscillates erratically during game play. It has been suggested that this oscillation in network latency, known as jitter, is much more destructive to the game playing experience than fixed delays (Delaney et al., 2004), such as those modeled in this study. Thus, future work must attempt to better understand the role of jitter on user experience in online gaming.

Chapter 3

Experiments 3.1 and 3.2

3.1 Experiment 3.1

The previous four experiments were designed to analyse the functional relationships between the variables of Relational Complexity and Network Latency, and to assess the effects of these variables on a number of measures. The analysis has suggested that; 1) the effects of complexity may be observed in only a narrow range of game types that are neither very high nor low in complexity, and 2) that network delay appears to interact in a very complex way with complexity. Indeed, it appears that both floor and ceiling effects may have been in operation in the previous four experiments.

Neither Experiment 2.1 nor 2.2 found a significant effect for Relational Complexity on game performance or on any of the subjective measures. In effect, as complexity increased across levels of the game, participants' performance in and experience of the game were unaffected, suggesting that the equivalence level was not any more difficult than the generalisation level. However, the level of Relational Complexity employed in Experiments 2.1 and 2.2 was relatively low, consisting of stimulus generalisation and one-node stimulus equivalence. Indeed, closer inspection of Correct Responses reveals that the majority of participants in all conditions in these experiments produced responding at a level that would constitute a passing criterion in a typical equivalence testing procedure. Thus, it appears that the level of correct responding was so high in both the generalisation and equivalence levels that it was impossible to distinguish these levels statistically. It may be reasonably argued, therefore, that participants did not experience significant challenge in playing these games. As such, the consistency in ratings of enjoyment, difficulty and frustration across levels is understandable. In simple terms, requiring participants to respond to stimulus

equivalence and stimulus generalisation relations across different levels of a game may not represent a sufficient manipulation of complexity to observe differences in effects across levels.

In Experiments 2.1 and 2.2, participants who played the game containing delays achieved significantly lower scores on that game. In addition, ratings of difficulty and frustration were significantly higher for participants who played a game containing simulated network delays than for those who played a game containing no delays. These findings suggest that the presence of simulated network delays appears highly destructive towards the game playing experience. Thus, it appears that while the level of complexity employed was too low to distinguish levels of varying complexity, network delay still had an effect on both subjective and behavioural measures.

As effects had been established for Network Latency in Experiments 2.1 and 2.2, Experiments 2.3 and 2.4 aimed to increase the level of relational responding required to play the game, in order to eliminate the ceiling effect observed on the complexity variable in Experiments 2.1 and 2.2, while also extending the analysis of the delay variable. Relational Complexity was manipulated across stimulus equivalence relations containing larger numbers of stimuli in experiments 2.3 and 2.4 than those employed in Experiments 2.1 and 2.2.

In Experiment 2.3, significant effects were found for Relational Complexity, in that more complex levels produced fewer Correct Responses and were rated as less positive than less complex levels. This finding strongly suggests that the complexity variable was manipulated correctly in this experiment, insofar as higher levels of complexity typically lead to lower scores. Importantly, Relational Complexity did significantly affect Positive Affect ratings in Experiment 2.3. Specifically, there was a significant trend of lower Positive Affect ratings for more complex levels of the game in comparison with the less complex levels in both experiments. Thus, these findings confirm that through manipulating the level of derived relational responding required in a game, the level of difficulty or challenge experienced by the players can be increased.

No effects were found for the Network Latency variable on Correct Responses in Experiment 2.3. An analysis of mean scores in this game suggested that a ceiling effect may have again been in operation, as the majority of participants produced scores that would constitute a pass on a typical equivalence testing phase (i.e., approximately 90% correct responding or higher). Thus, Experiment 2.4 was conducted in order to eliminate this ceiling effect, so that the effects of 0.5s and 1s delays on game play and experience could be examined.

The significant effect found for the variable of Relational Complexity on Correct Responses in Experiment 2.3 was not replicated in Experiment 2.4. In Experiment 2.4, participants were required to produce the relevant derived relational responses for the first time during the game phase (i.e.,

participants were not presented with a relational testing phase before playing the game). This procedure was conceived of as considerably more difficult than that employed in Experiment 2.3. That no significant difference was found between Correct Responses on the earlier levels of the game and those on the later levels may be explained by the relatively high demand placed on participants in Experiment 2.4 compared to Experiment 2.3. Essentially, a floor effect may have been observed on scores in Experiment 2.4. Indeed, the mean number of Correct Responses recorded across all participants in each condition in Experiment 2.4 is approximately that which would be expected by chance. When the raw data is examined, it is evident that only a small minority of participants in each condition reached a 90% correct criterion. Thus, it appears that correct responding was uniformly low and inconsistent across all levels in Experiment 2.4. In effect, it appears that while a ceiling effect may have existed in Experiments 2.1 and 2.2, where the level of relational responding required by the game was not difficult enough, the opposite may have been the case in Experiment 2.4. It appears that only Experiment 2.3 offered a sufficiently broad spectrum of difficulty levels to produce significantly different amounts of Correct Responses across levels and participants. Thus, it appears that the effects of complexity on scores achieved in a game may be observed in only a narrow range of game types that are neither very high nor low in complexity.

Relational Complexity did significantly affect Positive Affect ratings in Experiment 2.4. Specifically, there was a significant trend of lower Positive Affect ratings for more complex levels of the game in comparison with the less complex levels in both experiments. Thus, the finding from both Experiments 2.3 and 2.4, that participants prefer to engage in less challenging games may be of considerable interest to the computer games industry. Specifically, these findings challenge the conventional wisdom in the engineering literature that suggests a link between higher levels of challenge and enjoyment in games (Davis, Steury, & Pagulayan, 2005; Gingold, 2005; Kiili, 2005a; Malone, 1982; Morlock, Yando, & Nigolean, 1985).

In Experiments 2.3 and 2.4, participants were presented with a game containing either short 0.5s delays or long 1s delays. No significant difference was found between these groups, on any measure employed in the study. However, as there was no 'no delay' condition, the results of Experiments 2.3 and 2.4 do not suggest that Network Latency has no effect on participants' performances or enjoyment of games more generally. Rather, the results of all four studies combined may suggest that while the presence of delay in a game is destructive to the game playing experience, increasing the length of that delay beyond 0.5s does not have any further negative effects.

Of course, it could be argued that the levels of relational responding examined in the experiments conducted here were relatively low compared to those required in a large number of

modern games. The game format itself, involving a relatively static graphic environment and a relatively small number of monochrome stimuli, may have proved unstimulating for participants, who are familiar with games involving high quality 3D graphics and the simultaneous presence of multiple stimuli on-screen, augmented with hi-fidelity sound effects. It remains a possibility, therefore, that in a more engaging game environment, participants would prefer more complex games. Alternatively, it is possible that in a game that requires more complex relational responding, the higher levels of complexity would be more enjoyable to participants. In effect, it might not be that participants enjoy less relationally complex games across the board, but appear to do so only when the levels under examination are low and the game format is relatively unstimulating.

The suggestion that the previous games were not sufficiently complex is borne out by the high number of Correct Responses observed in all games at all levels, with the exception of Experiment 2.4, in which a floor effect was observed. Indeed, despite the time constraints employed in the game context, the majority of participants in Experiments 2.1, 2.2 and 2.3 actually reached a level of correct responding that would be deemed an acceptable pass rate for a standard stimulus equivalence testing phase. Most equivalence test procedures, including those employed in the previous experiments, involve multiple test exposures. That is, participants are typically exposed to an equivalence test block a number of times before they reach the passing criterion. Testing blocks are rarely passed on the first exposure. For example, Arntzen and Holth (1997) investigated a number of procedures for training equivalence relations to criterion. None of the participants who were trained in the linear method (i.e., the training procedure used in Experiments 2.1 - 2.4), reached a 90% passing criterion on their first exposure to an equivalence test. Arntzen and Holth (2000) report a similar study, in which only one of three participants trained in the linear method first reached a 90% passing criterion on their first exposure to an equivalence test. Adams, Fields and Verhave (1993) also report a study that examined different procedures for training equivalence relations. In a procedure that closely resembled that used in Experiments 2.1, 2.2, 2.3 and 2.4, participants, on average, required almost five exposures to an equivalence test before successfully reaching criterion on that test (see Fields, Adams, Verhave, & Newman, 1990; Fields, Reeve, Rosen, Varelas, & Adams, 1997; Saunders, Saunders, Williams, & Spradlin, 1993; Smeets & Barnes-Holmes, 2005, for further discussion). Thus, performances observed on even the most complex games presented in Experiments 2.1, 2.2 and 2.3 are particularly impressive considering the fact that each game level was presented only once. The important point, however, is that while subjects reported a preference for the less complex games in previous experiments, it does not follow that they are unable to play the more complex games as defined here.

It has been argued that the games presented in Experiments 2.1, 2.2 and 2.3 were not sufficiently complex to be enjoyable. This was supported by the relatively high scoring observed in these games. However, while the method of increasing complexity from Experiment 2.3 to Experiment 2.4 (i.e., removing the stimulus equivalence testing phase from the experimental procedure), did decrease the number of Correct Responses recorded by participants, it did not appear to affect their subjective ratings of enjoyment. Thus, while there may be some value in increasing the Relational Complexity presented in future games, in order to ascertain whether participants do indeed prefer simpler games, an alternative method of doing so than that employed in Chapter 2 may prove more fruitful.

Importantly, while group analyses of the data obtained in Experiments 2.1 - 2.4 produced ambiguous findings, single-subject analyses did little to clarify the relationship between the complexity and latency variables. There appear to be large differences in performances of participants within all groups, which suggests that the effects of Relational Complexity and Network Latency may be heavily dependent upon participants' particular histories. That is, it would appear that large differences in performance are observed across subjects exposed to similar latency and complexity levels under laboratory conditions. It may well be, therefore, that these variables operate in highly idiosyncratic ways not easily brought under experimental control for large groups. Moreover, the interactions between the variables of complexity and latency may also operate in idiosyncratic ways, given the personal and genetic histories of participants. Alternatively, the variables may be related to each other in a highly complex non-linear fashion. For example, it may be that the negative effects of Network Latency on game enjoyment would be less salient under a high level of Relational Complexity. Alternatively, it may be that under particularly low levels of Relational Complexity, some Network Latency might actually provide an enjoyable challenge to game players. However, such interactions have been very difficult to establish in the current research. Thus, it would appear that these two variables simply do not interact in a linear fashion and are also individually affected by idiosyncratic personal histories of participants.

Given the unclear nature of data gathered in the experiments conducted thus far, it is possible that there may be no consistent linear or easily ascertained relationship between the variables of Relational Complexity and Network Latency. Indeed, while it may be possible to disentangle the effects of these variables through refining the previous experiments, such an exercise may prove time consuming and fruitless. So, it may not be worthwhile to simply continue to pursue the latency - complexity interaction relationship until at least each of these variables has been examined more carefully independently of each other. Such a strategy will allow for a more measured and focused exploration of these variables as independent factors in game performance and enjoyment.

One possible method of manipulating Relational Complexity, which would allow for the functional definition of a number of different levels of complexity, is to require participants to engage in complex relational responding. The four previous experiments have utilised stimulus equivalence in manipulations of Relational Complexity. In Experiments 2.1 and 2.2, participants were required to respond to one node stimulus equivalence relations. In Experiments 2.3 and 2.4, participants were required to respond to both one node and three node equivalence relations. However, stimulus equivalence is not the only form of derived relational responding that may be employed in the manipulation of Relational Complexity. In fact, Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001a) treats stimulus equivalence as the simplest, or least complex, form of derived relational responding. Indeed, this is precisely the reason why equivalence relations were utilised in the initial studies. As the expected effect for Relational Complexity was not consistently observed with relatively simple equivalence responding, more complex forms of derived relational responding could now be utilized in order to fully investigate the effects of Relational Complexity on computer game experience and enjoyment.

A number of recent empirical studies (e.g., O’Hora, Roche, Barnes-Holmes, & Smeets, 2002; Roche, Linehan, Ward, Dymond, & Reheldt, 2004; Steele & Hayes, 1991; Whelan & Barnes-Holmes, 2004; Whelan, Barnes-Holmes, & Dymond, 2006) have demonstrated that participants can respond consistently in accordance with relations other than that of equivalence. All of these studies employed procedures analogous to those developed by Steele and Hayes (1991) and involve an initial pre-training stage, a relational training stage, followed by a relational testing phase. For example, Whelan and Barnes-Holmes (2004) demonstrated that participants can respond consistently in accordance with the relations of SAME and OPPOSITE. The procedure involved an initial pre-training stage, in which the functions of SAME and OPPOSITE were established for two arbitrary stimuli using a matching-to-sample type conditional discrimination training technique. Specifically, groups of visual stimuli, differing in only one non-arbitrary dimension were presented and participants were required to choose which of the comparison stimuli to match with the sample. Stimulus sets included a number of lines differing in length, a number of squares differing in size, and sets of dots differing in the number of dots included in the set. In the presence of the SAME contextual cue, choosing the comparison stimulus that was formally identical to the sample was reinforced, while choosing any other stimulus was punished. In the presence of the OPPOSITE contextual cue, choosing the stimulus which was the most different to the sample was reinforced, while any other choice was punished.

These contextual cues were then used to establish a relational network between five arbitrary

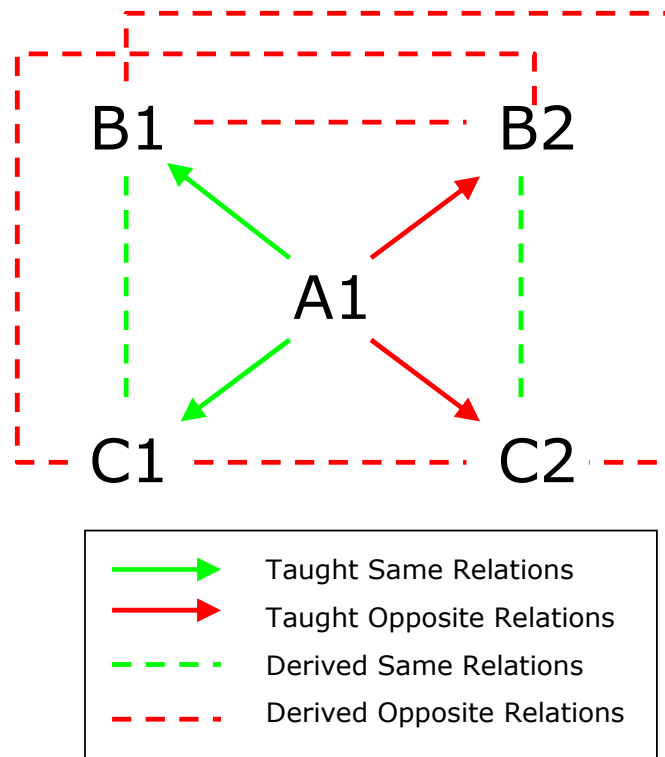


Figure 3.1: Representation of the relational network trained and tested in Whelan and Barnes-Holmes (2004).

visual stimuli, nominally referred to as A1, B1, B2, C1, C2 for the sake of clarity. A typical trial in this stage involved the presentation of a sample stimulus (A1) and two comparison stimuli (B1 and B2). Choosing B1 in the presence of A1 and the SAME contextual cue was reinforced, while choosing B2 in the presence of A1 and the OPPOSITE contextual cue was reinforced. In this method a complex network (as illustrated in figure 3.1) was established. A testing phase was then employed in which derived relations were tested. For example, as A1 was trained as opposite to both B2 and C2, choosing C2 in the presence of B2, and vice versa, in the presence of the SAME contextual cue, was deemed a correct response (as the opposite of an opposite entails a SAME relation).

Using a similar methodology, O’Hora, Roche, Barnes-Holmes and Smeets (2002) demonstrated that participants could produce stable responding in accordance with relations of MORE THAN, LESS THAN, SAME and OPPOSITE. Indeed, all three participants in Experiment 1 reported by O’Hora et al., produced stable responding in accordance with SAME and OPPOSITE relations on their first exposure to the relational testing phase. Stable responding in accordance with MORE THAN and LESS THAN was demonstrated by all three participants within three exposures to the

relational testing procedure.

Roche, Linehan, Ward, Dymond and Rehfeldt (2004) also reported on a study in which a number of participants demonstrated stable responding in accordance with the relations of SAME and OPPOSITE, again using a similar procedure to that developed by Steele and Hayes (1991). Whelan and Barnes-Holmes (2004) reported both stable responding in terms of SAME and OPPOSITE relations, and the transformation of consequential functions through these relations. Finally, Whelan, Barnes-Holmes and Dymond (2006) reported stable responding in accordance with relations of MORE THAN and LESS THAN, and the transformation of consequential functions through these relations. Thus, a large amount of empirical research has established that patterns of derived relational responding other than equivalence responding can be generated in the laboratory.

It must be noted at this point that some debate has occurred over the precise process involved in derived relational responding involving relations other than equivalence. Alternate accounts to that provided by RFT have been suggested by a number of researchers (see Galizio, 2004; McIlvane, 2003; Palmer, 2004; Tonneau, 2001b,a). However, an analysis of these protracted conceptual issues is beyond the scope of the current thesis. For the purposes of the current chapter (i.e., defining and manipulating Relational Complexity for the purposes of examining enjoyment in computer gaming) RFT presents a coherent, practical and applicable account of derived relational responding.

According to RFT, relations such as Opposite, More Than and Less Than require more complex relational activities than those observed in equivalence responding (Barnes & Hampson, 1993; O'Hara et al., 2002; Steele & Hayes, 1991; Whelan & Barnes-Holmes, 2004; Whelan et al., 2006). In order to illustrate this point, consider the three defining properties of derived relational responding; mutual entailment, combinatorial entailment and transformation of function.

Mutual Entailment: If a stimulus A bears a relationship to another stimulus B, then a further derived relation between B and A is mutually entailed. The type of relation mutually entailed depends upon the nature of the relation between A and B (Hayes, 1994). For instance, if the stimulus A bears an equivalence or SAME relation to the stimulus B, then the relation "B is the same as A" is mutually entailed. Thus, relations mutually entailed from equivalence relations are always equivalence relations. However, this is not always the case when alternative relations are trained. For example, if the stimulus A bears a relation of comparison to the stimulus B (e.g., A is MORE THAN B), then the relation "B is less than A" is mutually entailed. In other words the derived response may be functionally distinct from the trained response. The distinction is even more apparent when we consider combinatorial entailment.

Combinatorial Entailment: If a stimulus A bears a relation to another stimulus B, and B bears

a further relation to a stimulus C, then a relation between A and C is combinatorially entailed. The nature of the combinatorially entailed relation depends entirely on the nature of the trained relations. As with mutual entailment, when the relation trained is an equivalence relation, the relation entailed is always an equivalence relation. For example, if A is the SAME as B and B is the SAME as C, then A is the SAME as C. However, when relations other than those of equivalence are trained, quite complex relations may be entailed. For example, if A is the OPPOSITE of B and B is the OPPOSITE of C, then a SAME relation between A and C is derived by combinatorial entailment (i.e., A is the SAME as C). Importantly, RFT suggests that relational responding in accordance with a frame in which the derived response is functionally distinct from the trained response constitutes a more complex relational activity than responding in accordance with a frame in which the derived response is functionally identical to the trained response. Thus, responding to the opposite of an opposite would represent a more complex relational activity than an equivalence response.

Transformation of Function: If a stimulus A is related to another stimulus B, and A acquires a psychological function, then in the appropriate context the stimulus functions of B will be transformed in accordance with the A-B relation. For example, if A is MORE THAN B, and A elicits fear, then B will produce less fear than A.

Importantly, through the concept of complex derived relational responding, RFT provides us with a method of functionally defining game levels that are more or less complex than each other. For example, a level which requires players to make derived SAME responses would be defined as functionally simpler than a level which requires participants to make derived OPPOSITE responses. This approach should allow for the investigation of participants' enjoyment and experience across a number of levels that differ significantly in terms of Relational Complexity.

An additional aim of any further research should be to also examine the effects of Network Latency on game playing across an increased range of delays experienced by participants. For example, Experiments 2.1 and 2.2 involved exposing participants to either no delay or a small amount of delay, while Experiments 2.3 and 2.4 exposed participants to either a small or medium amount of delay. None of these foregoing experiments have examined a full range of latencies on participants' game playing performance or experience. As such, it has been difficult to provide a complete account of the effects of latency on participants' game playing behaviour based on the results of the previous experiments. In order to better understand the effects of different amounts of delay on participants' behaviour, a method must be devised whereby participants are exposed to a number of different delay conditions, spanning the whole spectrum from no delay, through medium amounts of delay, to a large amount of delay.

Another aim of any further research should be to model Network Latency in a more ecologically valid format than achieved previously. In the four experiments reported thus far, Network Latency was modeled as a fixed interval of either 0.5 seconds or 1 second. It may be argued that, given that participants could predict the length of each delay suffered, the delays could have been perceived as a particular challenge of the game, rather than a nuisance or problem with the game. In practice, Network Latency is rarely, if ever, predictable, and typically oscillates erratically during game play. It has been suggested that this oscillation in Network Latency, known as jitter, is much more destructive to the game playing experience than fixed delays (Delaney et al., 2004), such as those modeled in the current research thus far. Thus, the experiments to follow will attempt to address the effect of jitter, rather than mere delay, on user experience in online gaming.

A further issue worth considering is that the previous experiments depended on a group design to extract out the various effects of multiple variables simultaneously. This endeavour may have been fruitful if the variables had consistent linear effects and interacted with each other in linear ways. However, these variables do not appear to have linear effects on each other or on the measures employed. Thus, a very large number of participants would be required in order to establish clear statistical effects for the various manipulations attempted thus far. However, it would not be in line with behaviour analytic traditions to pursue an unbounded increase in subject numbers in order to attain effects through sheer volume of data. In fact, the opposite strategy is required. What is needed is a clear view of behavioural process at the level of individual subjects as well as at a group level. As such, what is needed now is a more considered analysis with very low numbers of participants in a context in which the prediction and influence over participants' tests performances takes priority over the statistical generation of significant relationships between variables.

Group designs that require multivariate statistical analyses have been criticised extensively within the field of behaviour analysis (e.g., Baron, 1999; Branch, 1999; Perone, 1999). Such designs rely on inference rather than induction to establish findings, and rely on the outcomes of statistical tests, rather than the need for control over a priori behaviours of interest, to indicate interesting avenues of research. In effect, psychologists who use group designs can easily mistake statistical significance for psychological significance. Over reliance on statistical effects over behavioural control misdirects the research agenda from effective interaction with our participants, to finding statistically significant effects for their own sake. But more importantly, one basic assumption of psychological research, which is often forgotten when group designs are employed, is that any effects observed at the group level should be demonstrable with a very low number of participants; ideally 1 (i.e., every member of that group should demonstrate the behaviour under investigation).

It is possible that statistical significance might be observed in enjoyment level differences across levels of a game. However, there may also be a relatively large overlap in responses to the questionnaire, without jeopardising the overall statistical effect due to the large number of participants. However, we should not be satisfied with this level of statistical significance as it leaves those few individuals who showed ambiguous effects or effects in the opposite direction unaccounted for. For instance, some participants may demonstrate large effects while others demonstrate small effects. A group-based analysis may miss this important feature of the data entirely. While many psychologists are unperturbed about the exceptions to their experimental manipulations, it has been central to the radical behavioural approach that these unexpected outcomes are the most interesting from a behavioural control point of view (see Sidman, 1960). While we may only aspire to achieving perfect levels of control in the current research, this should nevertheless remain the ultimate goal of an experimental analysis. In the experiments to follow more careful attention will be paid to establishing clear effects for the experimental variables across a range of individual participants. Consequently, statistical inference will not be used as the main tool for analysing outcomes of these experimental manipulations. Rather, a within-subjects design with a small number of participants will be employed. The two variables of interest will be investigated independently of each other.

Experiment 3.1 aimed to; 1) model a game involving derived relations that presented different levels of 'challenge'/ Relational Complexity to participants, and 2) measure the effect of these different levels of complexity on participants' performance and subjective ratings of enjoyment. Importantly, participants will be required to produce derived relational responses in accordance with the relational frames of SAME and OPPOSITE during the game phase. This represents a more complex level of relational responding than employed in previous experiments. Participants will first be exposed to a relational pre-training phase, designed to establish the functions of SAME and OPPOSITE for two arbitrary contextual cues. A relational training phase will then be presented. This phase was designed to create a complex network of relations between a number of arbitrary stimuli, using the contextual cues established in relational pre-training. Upon successful completion of the relational training and testing procedure, participants will be exposed to a computer-presented game, in which the game characters are those stimuli which were used in the relational training phase. Importantly, four levels will be presented, requiring four different types of derived relational responding. It is conceived that level 1, which requires no derived responding, will present considerably less of a challenge than level 4, which requires both SAME and OPPOSITE responding across inter-mixed trials. No manipulation of the Network Latency variable will be carried out.

Experiment 3.2 aimed to investigate the role of Network Latency on computer game playing.

The same relational pre-training and relational training procedure will be employed as that used in Experiment 3.1. However, the gaming phase differs considerably from Experiment 3.1, in that the level of derived relational responding required is held constant, whereas three different levels of Network Latency, ranging from no latency to a large amount of latency, are applied across three different games of the same type. In addition, Network Latency is simulated in a more ecologically valid method than previous studies in order to more accurately emulate jitter, rather than mere delay.

3.1.1 Method

3.1.1.1 Participants

Thirty-three participants (14 male, 19 female), all first year undergraduate students aged 18 - 35, were recruited for Experiment 3.1 through personal contacts, notice board advertisements and cold calling on individuals around the University campus. Participants were not paid for their participation in the study.

3.1.1.2 Materials

The experiment was conducted in a quiet room free of distraction. All stimuli were presented on a computer screen (resolution 1024 x 768) via Microsoft Visual Basic 6.0 software (c.f., Dixon & MacLin, 2003; Cabello, Barnes-Holmes, O'Hora, & Stewart, 2002) which also recorded the nature and timing of stimulus presentation and participant responses.

3.1.1.3 Design

The experiment employed a repeated measures design with one independent variable, Relational Complexity, which was manipulated at four different levels. All participants were exposed to four experimental conditions (Level 1 –Training, Level 2 –SAME, Level 3 –OPPOSITE, and Level 4 –Mixed SAME and OPPOSITE) in a counterbalanced order. There were five dependent measures; participants' score on each level of the game, and their subjectively rated level of both Positive and Negative Affect, as well as Competence and Impatience (see Kahneman, et al., 2004).

3.1.1.4 Procedure

The experiment was divided into three stages; (i) Relational Pre-training, (ii) Relational Training, and (iii) Gaming

(i) Relational Pre-training and Testing

Participants were first presented with relational pre-training, designed to establish the functions of SAME and OPPOSITE, respectively, for two arbitrary stimuli, which functioned as contextual cues for the entire experiment. Specifically, one of two colours (beige and lilac) was presented as background colour during all trials in this phase. Responding appropriately in the presence of each colour across a number of trials was reinforced. Importantly, the stimuli employed in the relational pre-training phase consisted of sets of similar stimuli that differed along one non-arbitrary continuum. For example, stimulus sets included three lines of varying length, three circles of varied circumference and three boxes of varied shading. In a SAME pre-training trial, matching stimuli that were most similar to each other was reinforced, while in an OPPOSITE pre-training trial, matching stimuli that were most different to each other was reinforced. The relational pre-training stage consisted of five training blocks, followed by a testing block, all of which had to be passed to a pre-defined criterion in order for the participant to progress to the relational training stage.

Each training trial during the relational pre-training phase consisted of the following sequence of events. Firstly, the background colour of the trial was presented, and remained on- screen for the duration of the trial. On a SAME trial a beige background was presented, whereas a lilac background was presented on an OPPOSITE trial. After an interval of 1000ms, the sample stimulus was presented at the top of the screen, followed 1000ms later by the three comparison stimuli at the bottom of the screen. The three comparison stimuli were spread across the bottom of the screen so that one was presented in the bottom left hand corner, one was presented in the bottom right hand corner and the remaining stimulus was presented centrally at the bottom of the screen. The positions of these comparison stimuli was randomized across trials. The sample and three comparison stimuli remained on the screen until the subject made a response. Once a response was made by the participant, the screen cleared and feedback was presented to participants on a blank white screen. Specifically, if the response was defined by the experimenter as correct, the word, "Correct," was presented in the middle of the screen, whereas if the response was incorrect, the word, "Wrong," was presented. Feedback was presented in black 36 point font and remained on-screen for 2000ms before the screen was cleared and the next trial was presented.

Pre-training consisted of five training blocks followed by a testing block. The first two pre-training blocks were designed to establish the function of SAME for the beige background colour. The following two blocks were designed to establish the function of OPPOSITE for the lilac background colour. The final pre-training block involved the presentation of both SAME and OPPOSITE tasks intermixed in a quasi-random order.

In pre-training blocks one and two, choosing the comparison stimulus which was most similar to the sample was reinforced. Pre-training block one involved the presentation of a set of hexagons which differed along the dimension of size. Pre-training block two consisted of the presentation of a set of stimuli which contained different amounts of triangles of identical size. All pre-training blocks one and two consisted of twelve trials, eleven of which had to be responded to correctly in order for the participant to progress to the next stage. If a participant failed any one training block four times, the experiment was terminated and the participant was debriefed.

Pre-training blocks three and four involved the presentation of exactly the same sample and comparison stimuli as training blocks one and two, respectively. However, unlike in pre-training blocks one and two, these stimuli were presented in the presence of the OPPOSITE contextual cue (the lilac background colour). Choosing the comparison stimulus that was most different to the sample stimulus was reinforced. If a participant failed any one training block four times, the experiment was terminated and the participant was debriefed.

Pre-training block five required participants to respond correctly in the presence of both the SAME and OPPOSITE contextual cues on quasi-randomly alternating trials. A novel stimulus set consisting of black lines of three different lengths was employed in this stage, in order to ensure that participants were responding functionally in accordance with the SAME and OPPOSITE cues, and not in terms of any particular topographic features of the stimuli used in the previous stages. There were twenty four trials in total; twelve SAME trials and twelve OPPOSITE trials, presented in a quasi-random order. The participant was required to reach a criterion of 23/24 Correct Responses in order to progress to the testing phase. Responding in accordance with the contextual cue presented was reinforced. For example, in the presence of the SAME contextual cue, choosing the comparison stimulus that most closely resembled the sample stimulus was reinforced. Participants were required to repeat this stage indefinitely until they had met the response criterion, or thirty minutes had passed, whichever came first.

Relational pre-training testing consisted of a block of twenty four trials. As in pre-training block five, participants were required to respond correctly in the presence of both contextual cues on quasi-randomly alternating trials. However, no feedback was presented during a testing block. That is, after a response was made, the computer programme proceeded directly to the inter-trial interval. Two novel stimulus sets (three black squares of different size, and three squares of similar size but different colour) were employed in the first testing block. If participants did not pass the pre-training test on their first attempt, they were presented with another testing block, which contained one stimulus set from the first testing block and one novel stimulus set (groups of different amounts of

arrows). If participants did not reach criterion on the second testing block, they were returned to pre-training block five. After successfully passing pre-training block five, participants were presented with a third testing block. This block consisted of the stimulus set that was introduced for testing block two, plus another novel stimulus set (three faces of differing size). If participants did not pass the third testing block, they were presented with one final testing block. Once again, the stimulus set introduced for the previous stage was presented, along with a novel set (in this case a set of white cubes of differing size). If participants did not pass the pre-training testing phase on their fourth exposure, the session was terminated, participants were thanked and de-briefed. Once any relational pre-training testing block was passed, testing was considered to be complete and participants were immediately presented with the instructions for the relational training stage.

(ii) Relational Training and Testing

Relational Training was designed to create a complex network of relations between a number of arbitrary stimuli using the contextual cues established in relational pre-training. The stimuli used in relational training were five coloured 2D shapes; a yellow pentagon (A1), an orange arrow (B1), a red circle (B2), a blue triangle (C1) and a blue cross (C2). Relational training was conducted via three training blocks, followed by a testing block. See Figure 3.2 for an illustration of the relational network trained and tested in the current study.

All relational training blocks consisted of a matching-to-sample procedure, in which participants were instructed to choose which of two comparison stimuli ‘goes with’ a sample stimulus, in the presence of the contextual cues established in relational pre-training. Each task in relational training consisted of a similar procedure to that used in relational pre-training. First, a background colour was presented, followed 1000ms later by the sample stimulus, which in turn was followed 1000ms later by the comparison stimuli. However, unlike in pre-training, only two comparison stimuli were presented. Once participants made a response, the screen was cleared and feedback of, “Correct,” or, “Wrong,” was presented on-screen for 2000ms. When the 2000ms had elapsed, the screen was cleared and remained blank for a 1000ms inter-trial interval, following which the next trial was presented.

In relational training block one, participants were presented with the A1 stimulus as sample and the B1 and B2 stimuli as comparisons, in the presence of either the SAME or OPPOSITE contextual cues established in relational pre-training. The position of each comparison stimulus was counterbalanced across trials. Choosing B1 in the presence of A1 and the SAME contextual cue was reinforced, as was choosing B2 in the presence of A1 and the OPPOSITE contextual. There were twelve trials in this training block, eleven of which had to be responded to correctly before the

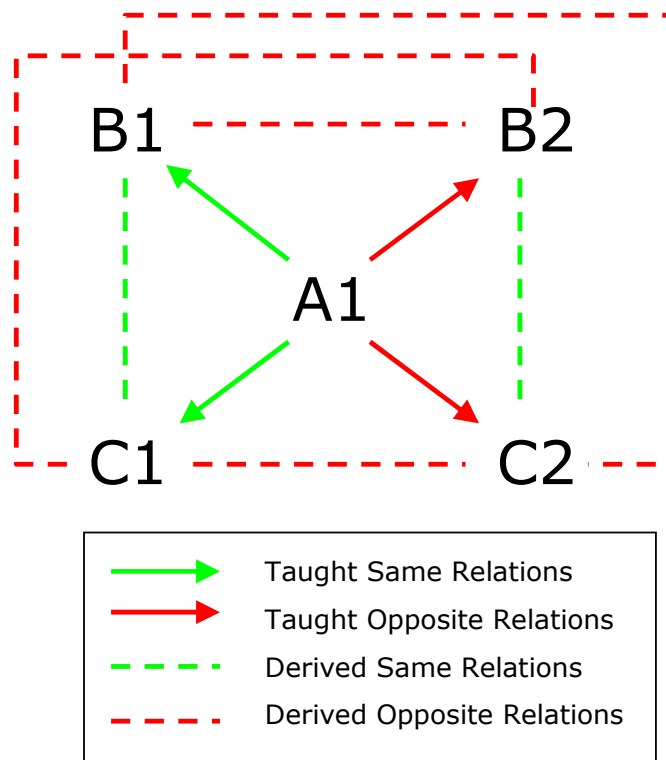


Figure 3.2: Representation of the relational network trained and tested in Experiment 3.1.

participant could progress to the following block.

Relational training block two was similar to block one, with the exception that the C1 and C2 stimuli were presented as comparisons. Choosing C1 in the presence of A1 and the SAME contextual cue was reinforced, while choosing C2 in the presence of A1 and the OPPOSITE contextual cue was reinforced.

Training block three involved the presentation of the A1 stimulus as sample and alternated randomly between the 'B' and 'C' stimuli as comparisons. Additionally, the SAME or OPPOSITE contextual cues were alternated quasi-randomly across all trials. There were sixteen trials in total during this training block, fifteen of which had to be responded to correctly before the participant could proceed to the testing phase. The participant received this training block repeatedly until they reached that criterion.

The relational testing phase probed for eight derived relations; SAME/B1-C1, SAME/C1-B1, SAME/B2-C2, SAME/C2-B2, OPPOSITE/B1-C2, OPPOSITE/C2-B1, OPPOSITE/B2-C1, and OPPOSITE/C1-B2. There were thirty two trials in total in the testing stage, representing four exposures to each trial type. Participants were required to respond correctly on thirty one trials in a testing block in order to progress to the game stage. If this criterion was not met, participants were returned to relational training block three, which they were required to pass before being presented with relational testing once more. The relational testing block was presented a maximum of four times. If participants did not pass relational testing within four attempts, the experiment was terminated and the participant was debriefed.

(ii) Gaming

Upon successful completion of the relational training and testing procedure, participants were exposed to a computer-presented game. The game consisted of four levels. Level 1 was a training level in which participants learned how to interact with the interface. Level 2 was a level in which participants were required to demonstrate derived relational responding in the presence of the SAME contextual cue. Level 3 was a level in which participants were required to demonstrate derived relational responding in the presence of the OPPOSITE contextual cue. Level 4 required participants to produce derived relational performances in the presence of quasi-randomly alternating SAME and OPPOSITE contextual cues. Importantly, the order in which these levels were presented was counterbalanced across participants. After each level a twelve-item questionnaire was presented, based on the DRM (Kahneman, et al., 2004) which measured two constructs; Positive Affect and Negative Affect, as well as ratings for Competence and Impatience.

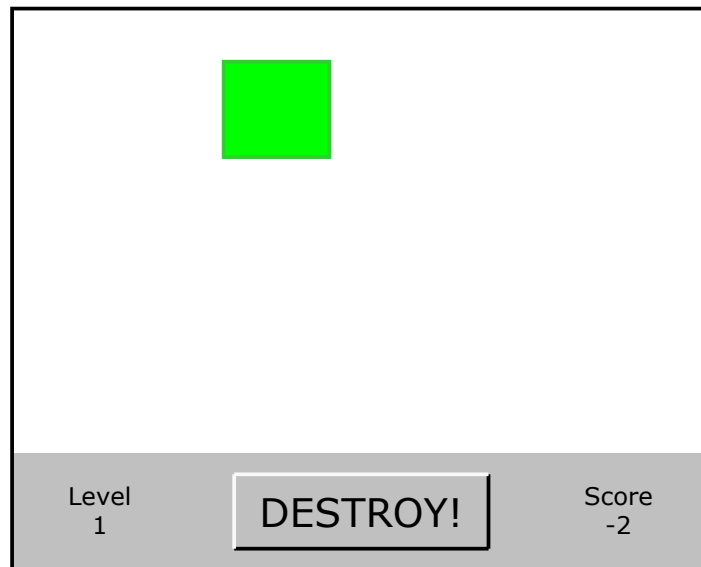


Figure 3.3: A screenshot of the user interface employed in the game phase of Experiment 3.1

As illustrated in Figure 3.3, the user interface involved a control panel at the bottom of the screen, with the current game level displayed in the left hand corner, the participant's score presented in the right hand corner and a button labelled 'DESTROY!' in the centre of the panel. Game characters were presented in twenty four possible locations in the area above the control panel.

A typical trial in the game phase consisted of the following sequence of events. A character was presented in one of twenty four possible locations on-screen. This character increased in size by 25% every 250ms in order to simulate approaching the participant. If the participant did not respond within 2000ms, the character disappeared from the screen and a point was deducted from the participant's score displayed in the bottom right hand corner of the screen. There were two responses available to participants on all trials, a save response, defined as clicking on the game character with the mouse pointer; and a destroy response, defined as clicking on the button labeled 'DESTROY!!!' on the control bar. When either a save or destroy response was recorded, the character displayed was removed from the screen, the destroy button became grey, a message displaying the phrase 'Response Detected' was presented beside the score indicator, and the score displayed in the bottom right hand corner of the screen was adjusted accordingly. If the response was correct, the score was increased by 1. If the response was incorrect, or no response was recorded, the score was reduced by 1. An inter-trial interval of 500ms was presented, after which the destroy button returned to its original appearance and the next character was presented.

Level 1, hereafter referred to as the Training Game, involved the presentation of the B1 and B2

stimuli from the relational training stage as game characters. Training was conducted in the presence of the SAME contextual cue (i.e., the beige background colour). Participants earned points for saving the B1 stimulus and destroying the B2 stimulus. Trials continued ad infinitum until a participant produced twelve consecutive Correct Responses.

Level 2 hereafter referred to as the SAME Game, involved the presentation of the C1 and C2 stimuli as game characters in the presence of the SAME contextual cue. Participants earned points for responding in accordance with derived SAME relations. For example, as points were gained in Level 1 for saving the B1 stimulus, and the C1 stimulus participated in a SAME relation with the B1 stimulus, points were gained in Level 2 for saving the C1 stimulus. Similarly, just as points were gained in Level 1 for destroying the B2 stimulus, points were gained in Level 3 for destroying the C2 stimulus. Crucially, scores were not displayed to participants during game play, because the presence of a score indicator on a trial-by-trial basis would have rendered Level 2 a training rather than a testing phase. That is, trial-by-trial feedback would quickly bring participants' responses under the direct contingency control of the displayed score. In effect, Level 2 would not function as a test for the participants' ability to engage in derived relational responding, but simply reflect participants' ability to respond in whatever way produced a score on each trial. Level 2 ended once 48 trials had been completed, regardless of the score achieved by participants. The score for this level was not presented to the participant until the entire game was completed.

Level 3, hereafter referred to as the OPPOSITE Game, involved the presentation of the C1 and C2 stimuli as game characters in the presence of the OPPOSITE contextual cue. As such, participants earned points for producing a response consistent with derived OPPOSITE relations. For example, as points were gained in Level 1 for saving the B1 stimulus, and the C2 stimulus participated in an OPPOSITE relation with the B1 stimulus, points were gained in Level 3 for saving the C2 stimulus (i.e., B1 is the opposite of C2). Likewise, points were gained for destroying the C1 stimulus, which participated in an OPPOSITE relation with the B2 stimulus. This level was employed as a more complex than Level 2. This level ended once 48 trials had been completed regardless of the participants' responses on each trial. Once again, the score for this level was not presented to the participant until the entire game was completed.

Level 4, hereafter referred to as the Mixed SAME and OPPOSITE Game, was essentially a combination of Levels 2 and 3. The C1 and C2 stimuli were presented as game characters in the presence of quasi-randomly alternating SAME and OPPOSITE contextual cues. Points were earned for making a response appropriate to the contextual cue presented. For example, points were earned for saving the C1 stimulus in the presence of the SAME contextual cue, whereas points were earned

for destroying the C1 stimulus in the presence of the OPPOSITE contextual cue. Importantly this level was employed as a more complex level than levels 2 and 3. Level 4 ended once 48 trials had been completed regardless of participants' responses. When all four levels were completed, the score achieved on each level was presented.

A twelve-item questionnaire was presented to participants after the completion of each individual level in the game as a subjective measure of both Positive and Negative Affect towards that level. The questionnaire used forms part of the Day Reconstruction Method (DRM; Kahnemann et al., 2004).

3.1.2 Results and Discussion

3.1.2.1 Relational Pre-Training

Table 3.1 shows the number of attempts each participant took to pass each stage in the Relational Pre-Training and Testing procedure. Note that an 'x' in a test block represents a failure to pass the test on that attempt. Once one training block had been passed, participants advanced to the following stage and were not exposed to any further training or testing. All participants passed the first two training blocks. One participant (P28) failed to pass the third training block after a large amount of attempts at doing so. Two further participants (P1 and P3) failed to pass the fifth training block. All three were thanked for their participation in the experiment and debriefed. Of the remaining thirty participants, twenty six passed the relational pre-training test on their first exposure, while four participants required two exposures to the testing phase. Thus, thirty participants advanced to the Relational Training phase.

3.1.2.2 Relational Training and Testing

Table 3.2 shows the number of attempts each participant took to pass each stage in the Relational Training and Testing procedure. Note that an 'x' in a test block represents a failure to pass the test on that attempt. Once one training block had been passed, participants advanced to the following stage and were not exposed to any further training or testing. All participants passed Relational Training phases one, two and three within three attempts at each phase. Eleven participants passed the relational testing phase on their first exposure. Of those who did not, all passed the following training phase within two exposures. Three participants passed relational testing on their second exposure. All sixteen remaining participants passed the following training phase on their first exposure. None of the remaining participants passed the relational test on their

Table 3.1: Number of attempts to pass each stage in the Relational Pre-Training and Testing procedure.

Participant	Train1	Train2	Train3	Train4	Train5	Test1	Test2	Train5	Test3	Test4
1	1	1	2	1	24					
2	1	1	2	1	2	x	1			
3	1	1	2	1	27					
4	1	1	2	1	1	1				
5	1	1	2	1	1	1				
6	1	1	2	1	4	1				
7	1	1	2	1	3	1				
8	1	1	4	1	3	x	1			
9	1	1	1	1	1	1				
10	1	1	1	1	1	1				
11	2	1	2	1	2	1				
12	1	1	1	1	2	1				
13	1	1	2	1	2	1				
14	1	1	2	2	2	1				
15	1	1	2	1	1	1				
16	2	1	2	1	3	1				
17	1	1	1	1	1	1				
18	1	1	2	1	2	1				
19	1	1	2	1	1	1				
20	1	1	2	1	1	1				
21	1	1	2	1	4	1				
22	1	1	2	1	1	1				
23	1	1	1	1	3	1				
24	1	1	2	1	2	1				
25	1	1	5	2	3	x	1			
26	1	1	42	1	4	1				
27	1	1	2	1	2	1				
28	1	1	49							
29	1	1	2	1	2	x	1			
30	1	1	1	1	1	1				
31	1	1	9	1	1	1				
32	1	1	2	1	1	1				
33	1	1	6	2	2	1				

Table 3.2: Number of attempts to pass each stage in the Relational Training and Testing procedure

Participant	Train1	Train2	Train3	Test1	Train3	Test2	Train3	Test3	Train3	Test4
2	1	1	1	1						
4	1	1	2	x	1	x	1	x	1	x
5	1	1	1	1						
6	1	1	1	1						
7	3	1	2	x	1	x	1	x	1	x
8	2	2	2	x	1	x	1	x	1	x
9	1	1	1	1						
10	1	1	1	x	1	1				
11	1	1	1	x	1	1				
12	1	1	1	1						
13	2	2	3	x	1	x	1	x	1	x
14	1	1	1	x	2	x	1	x	1	x
15	1	1	1	x	1	1				
16	3	1	1	1						
17	1	1	1	x	1	x	1	x	1	x
18	1	2	1	1						
19	1	1	1	x	1	x	1	x	1	x
20	2	1	1	x	1	x	1	x	1	x
21	1	1	1	1						
22	2	2	1	x	2	x	1	x	1	x
23	1	1	1	x	1	x	1	x	1	x
24	1	1	1	x	1	x	1	x	1	1
25	1	1	1	x	1	x	1	x	2	x
26	1	1	1	1						
27	2	1	1	x	1	x	1	x	1	x
29	1	1	1	1						
30	1	1	1	x	1	x	1	x	1	1
31	1	3	2	x	1	x	1	x	1	x
32	1	2	1	x	1	x	1	x	1	x
33	1	1	1	1						

third exposure. However, after passing the following training phase, a further two participants passed the testing phase on their fourth exposure. A total of fourteen participants did not pass the relational testing phase within four attempts, so their participation in the experiment was terminated. Thus, sixteen participants advanced to the gaming phase.

3.1.2.3 Gaming

Level 1 of the gaming phase was a Training Game in which participants learned how to interact with the user interface. All participants passed this Training Game and advanced on to the remaining games. Table 3.3 presents participants' scores across the SAME, OPPOSITE and Mixed SAME and OPPOSITE Games. Producing Correct Responses on 90% of all trials constitutes a

typical pass rate for a relational testing phase (Adams, Fields and Verhave, 1993; Arntzen & Holth, 1997; Arntzen & Holth, 2000;). Applying a similar criterion to each game may help to illuminate the difficulty experienced by participants in responding in accordance with previously established derived relations during those games. Fourteen of the sixteen participants produced Correct Responses on 90% or more of trials in the SAME Game. Twelve participants produced Correct Responses on 90% or more of trials in the OPPOSITE Game. Only five of the sixteen participants reached a 90% criterion in the Mixed SAME and OPPOSITE Game. Thus, it appears that more participants had difficulty responding in accordance with previously established derived relations during the Mixed SAME and OPPOSITE Game than in the SAME or OPPOSITE Games. There were four distinct patterns of responding displayed by participants across games during the gaming phase.

Ten of the sixteen participants produced more Correct Responses in the SAME Game than the OPPOSITE Game and also produced more Correct Responses during the OPPOSITE Game than during the Mixed SAME and OPPOSITE Game. This finding suggests that most (i.e. 10/16) participants found the Mixed SAME and OPPOSITE Game (i.e., the most relationally complex game) the most difficult, while the SAME level (the least relationally complex game) the least difficult.

Two participants (P6 and P29) produced a pattern of dramatically fewer Correct Responses on the OPPOSITE and the Mixed SAME and OPPOSITE Games, in comparison with the SAME Game. This strongly indicates that while responding in the SAME Game was under contextual control, responding during the latter two phases was not. That is, during the OPPOSITE phase participants were responding as if the SAME cue was present and therefore displayed perfect counter-control (i.e., 100% incorrect responses). It is difficult to imagine another systematic pattern of responding that would account for this effect. Similarly, during the Mixed SAME and OPPOSITE Game these participants typically produced correct responding when the SAME cue was present but further counter control (or absence of control by the opposite cue) during the OPPOSITE tasks. See Table 3.3 for a breakdown of Correct Responses in the presence of each contextual cue.

The performances of P18 and P24 are somewhat unexpected insofar as both of these participants passed the relational testing phase with 96% accuracy, but did not display consistently correct derived relational responding in any game presented. However, the current game established contingencies that competed with the purely derived relational responding context employed in that testing phase. That is, the game involved a time demand and as such required fluency as opposed to accuracy alone. This might be expected to lead to a deterioration of performance, particularly on

Table 3.3: Number of Correct Responses in the presence of each relational cue during the Game Phase

Participant	Same	Opposite	Mixed Same and Opposite	
			Same	Opposite
P2	48	48	22	23
P5	45	44	18	16
P6	47	0	14	7
P9	46	48	24	24
P10	45	48	24	23
P11	47	46	22	24
P12	47	47	22	24
P15	48	47	22	19
P16	45	44	21	19
P18	32	10	11	18
P21	48	48	14	7
P24	35	15	14	13
P26	48	46	22	15
P29	48	0	21	1
P30	48	48	14	14
P33	48	48	20	17

more difficult (i.e., relationally complex) games, as was observed here. Indeed, it could be argued, this increased response demand also accounts for the poor performances across all levels of the game displayed by P18 and P24.

P6, P29, P18 and P24 were the weakest gamers, in that the control exerted by the contextual cues in the game context was weakest for these participants. P6 and P29 displayed perfect counter-control in the OPPOSITE Game and poor responding to the opposite cue in the Mixed SAME and OPPOSITE Game, while P18 and P24 did not display consistently correct responding in the presence of either SAME or OPPOSITE cues in any game. Interestingly, three of the four weakest gamers (P6, P18 and P29) were among the strongest learners in the relational training and testing phase. Specifically, none of these participants required more than two exposures to any relational training block and more than one exposure to a relational testing block to pass those stages. Thus, it may be argued that these participants were exposed to less training and testing than other participants, and therefore the trained relations may have been less robust. In other words, participants who failed a number of relational testing phases would have received more training, thus would have a longer and more established history with the stimuli employed in the game. However, a number of participants (i.e., P2, P9, P12 & P16) produced similarly good performances in the relational training and testing phases before also producing consistently Correct Responses across all games. In addition, P24 performed very poorly in the gaming phase, but did receive a large number

of training phases, due to three relational test failures.

The foregoing issues are beyond exploration in this context but should serve as suggestions for where to start examining further variables that may affect the performance during the games. For instance, perhaps there are certain experimental conditions or personal histories that make the transfer of response functions from the B to the C stimuli more likely. Only further experimentation will resolve these issues. Importantly, however, 10 of the 16 participants did score lower in the Mixed SAME and OPPOSITE Game than the OPPOSITE Game and also scored lower in the OPPOSITE Game than in the SAME Game. In addition, 14 of the 16 participants produced less Correct Responses in the Mixed SAME and OPPOSITE Game than in the SAME Game. Thus, it appears that increased complexity as was defined here did in fact correlate with lower scores.

One participant (P9) displayed a consistent increase in scores across the SAME, OPPOSITE and Mixed SAME and OPPOSITE Games. This pattern was contrary to what would be predicted by an RFT analysis. However, it is important to note that this participant scored above 90% correct in all three games, and thus a ceiling effect may have been in operation. Thus, this participant could not reasonably be said to have experienced difficulty in any of the levels. Figure 3.4 illustrates graphically each participants' pattern of correct responding across the three games.

An analysis of mean Correct Responses for all participants, as presented in Table 3.4, raises an interesting but common statistical conundrum. The scores of a small minority of participants who displayed a different pattern of responding than the rest of the group are affecting the means in such a way that the typical pattern of responding of the majority of participants is obscured. Specifically, means for the OPPOSITE and Mixed SAME and OPPOSITE Games were very similar to each other, while quite different from the SAME Game. However, from a single subject analysis of the raw data, as presented graphically in figure 3.4, it appears that four participants' scores dropped considerably across the SAME and OPPOSITE Games, and increased across the OPPOSITE and Mixed SAME and OPPOSITE Games. In contrast, ten of the remaining participants' scores dropped consistently, but less dramatically, across the three games. In effect, these two patterns lead to unrepresentative mean scores when combined. The four participants who produced a very low number of Correct Responses in the OPPOSITE Game have dragged the mean scores down for this level, so that the mean score appears similar to the mean score for the Mixed SAME and OPPOSITE Game.

Figure 3.6 presents participants' ratings of Positive Affect, Competence, Negative Affect and Impatience across the SAME, OPPOSITE and Mixed SAME and OPPOSITE Games. Seven out of sixteen participants rated the OPPOSITE Game as less positive than the SAME Game, while ten

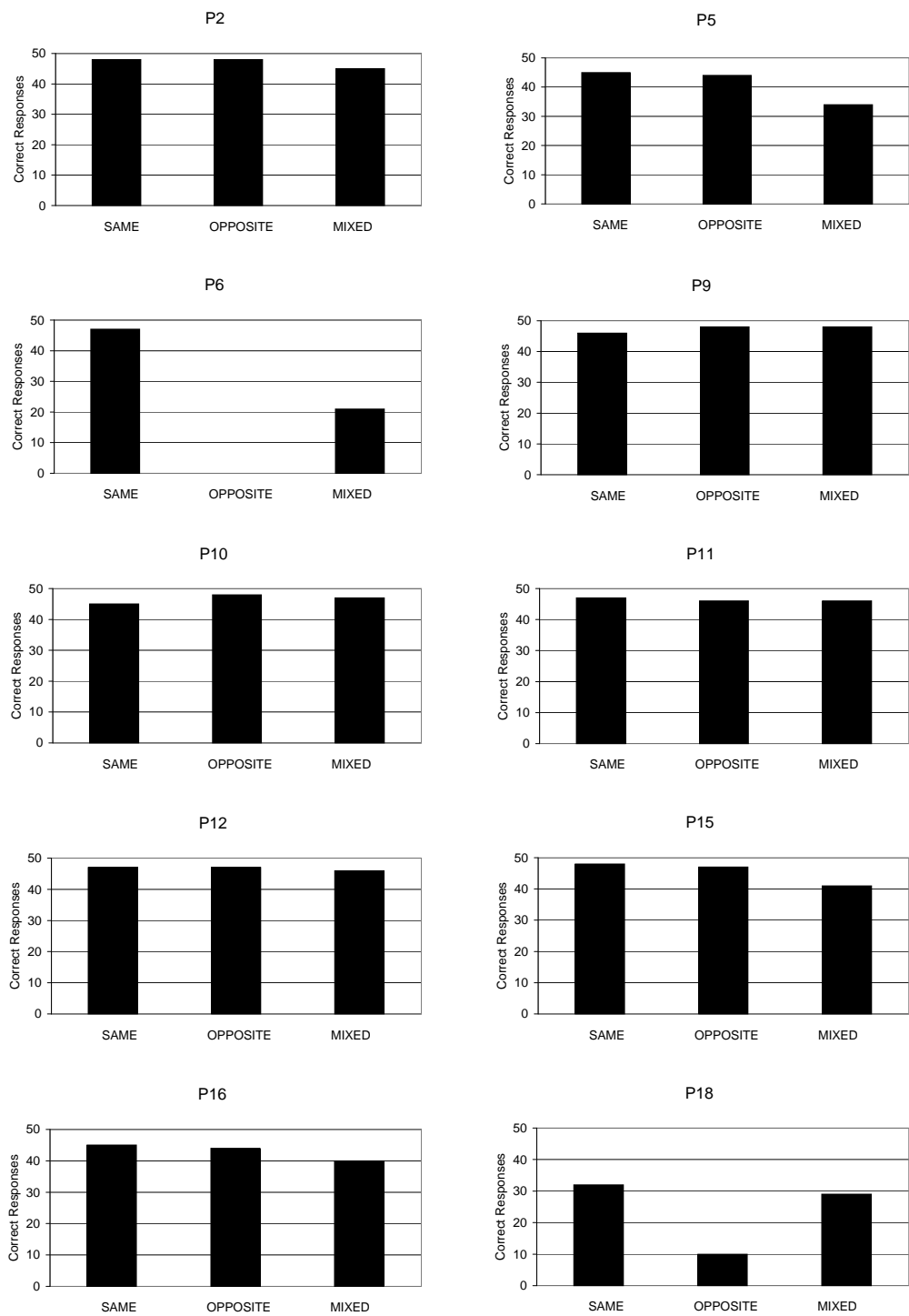


Figure 3.4: Correct Responses across the SAME, OPPOSITE and Mixed SAME and OPPOSITE Levels for each participant

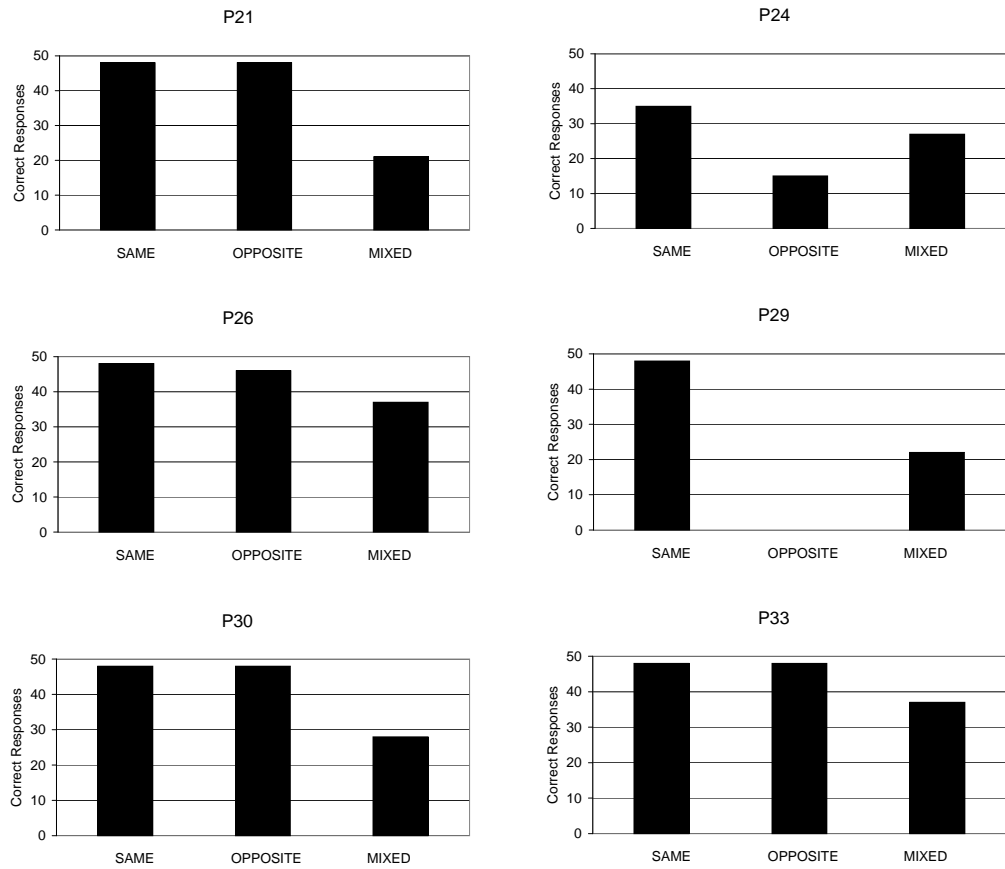


Figure 3.5: Figure 3.4 continued

Table 3.4: Mean Correct Responses across all four levels of the game (n=16)

	SAME	OPPOSITE	MIXED
MEAN Correct Responses	45.3	36.6	35.5

Table 3.5: Mean ratings of Positive Affect, Competence, Negative Affect and Impatience across all four levels in the gaming phase (n=16)

	Training	SAME	OPPOSITE	MIXED
Positive Affect	7.1875	6.75	6.5	4.3125
Competence	3.375	3.0625	3	2.0625
Negative Affect	6.75	5.5	6.375	9.6875
Impatience	2.0625	1.9375	2.3125	3.1875

participants rated the Mixed SAME and OPPOSITE Game as less positive than the OPPOSITE Game. Thus, there appears to be a pattern of the more complex levels being rated as lower in Positive Affect than the less complex levels.

Seven participants rated the OPPOSITE Game as lower for Competence than the SAME Game, while nine participants rated the Mixed SAME and OPPOSITE Game as lower for Competence than the OPPOSITE Game. Thus, mirroring Positive Affect scores, there appears to be a pattern in which the more complex levels are rated as lower in Competence than the less complex levels.

Seven participants rated the SAME Game as higher in Negative Affect than the Training Level, six participants rated the OPPOSITE Game as higher in Negative Affect than the SAME Game, while ten participants rated the Mixed SAME and OPPOSITE Game as higher in Negative Affect than the OPPOSITE Game. Thus, there appears to be a pattern wherein the more complex games are rated as higher in Negative Affect than the less complex games.

Seven participants rated the OPPOSITE Game as higher in Impatience scores than the SAME Game, while ten participants rated the Mixed SAME and OPPOSITE Game as higher in Impatience than the OPPOSITE Game. This mirrors Negative Affect scores, and there appears to be a pattern of higher Impatience scores reported for the more complex games.

In summary, participants appeared to rate the more complex games as lower in Positive Affect and Competence than the less complex games, while they also rated the more complex games as higher in Negative Affect and Impatience. While this pattern is not evident across all participants and across all games, it is evident across the majority of participants in most instances and is found on all four subjective measures employed. Thus, it appears that participants generally preferred the less relationally complex games over the more relationally complex games.

Table 3.5 presents mean ratings of Positive Affect, Competence, Negative Affect and Impatience across all four levels in the gaming phase. Mean Positive Affect ratings were higher in the relationally less complex games. Specifically, ratings of Positive Affect decreased linearly as Relational Complexity increased. A similar pattern was observed for mean ratings of Competence.

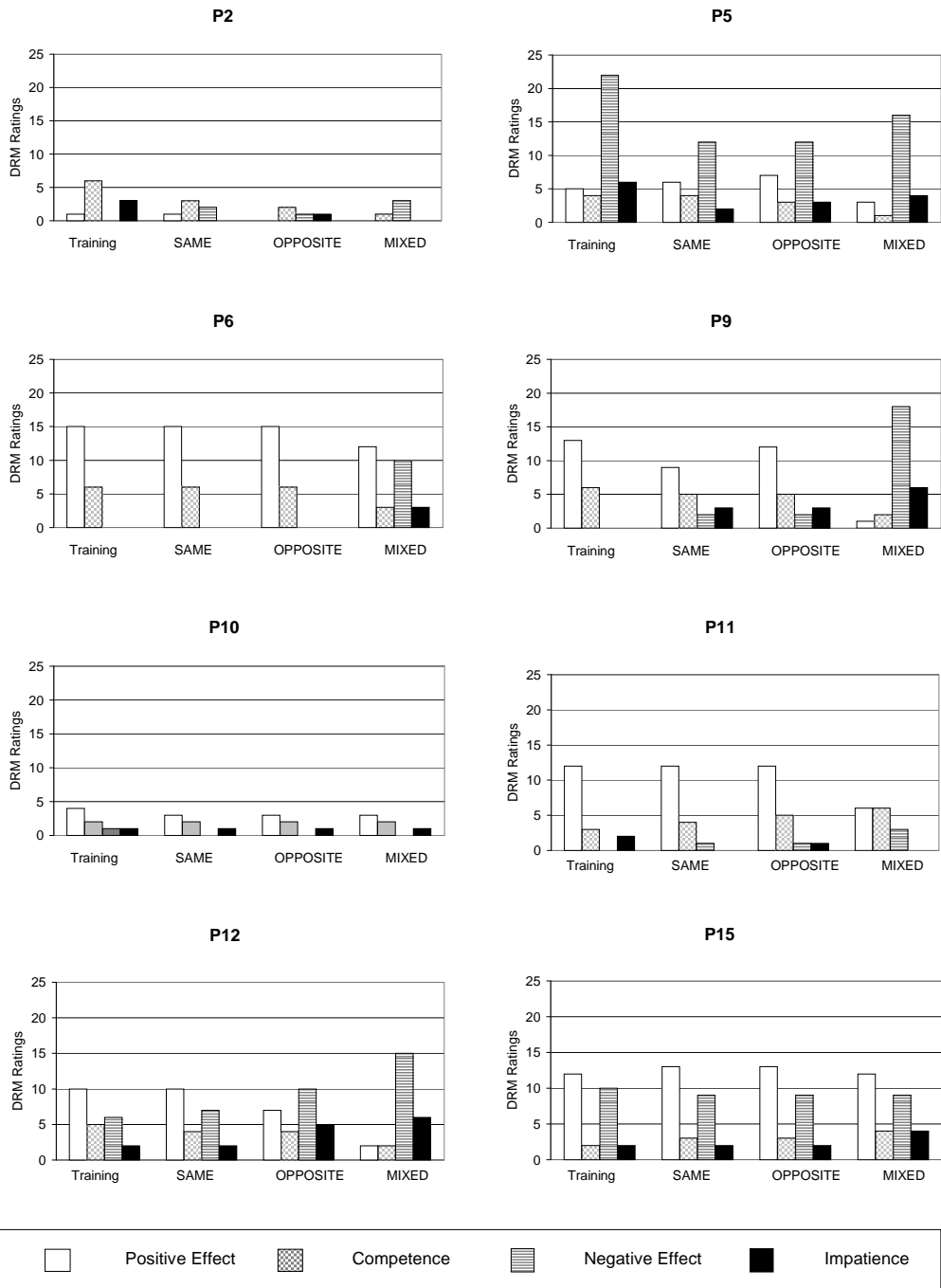


Figure 3.6: Ratings of Positive Affect, Competence, Negative Affect and Impatience across the SAME, OPPOSITE and Mixed SAME and OPPOSITE Levels for each participant

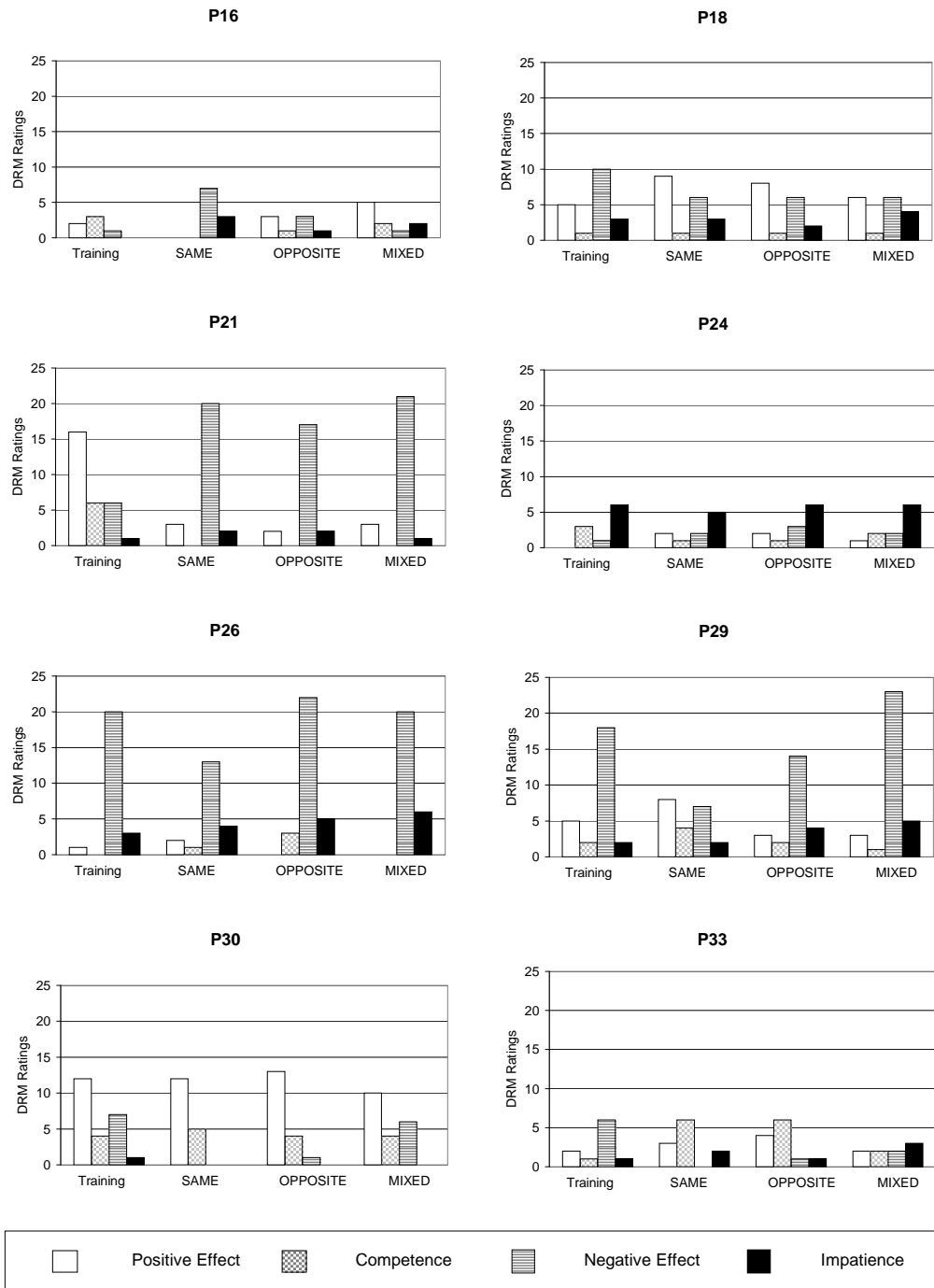


Figure 3.7: Figure 3.6 continued

Table 3.6: Results of a one-way Repeated Measures ANOVA examining the effects of Relational Complexity on Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=16)

	Wilks' Lambda	F Value	P Value	Eta Squated
Correct Responses	0.494	7.164	0.007	0.506
Positive Affect	0.598	2.916	0.074	0.402
Competence	0.673	2.110	0.148	0.327
Negative Affect	0.647	2.367	0.118	0.353
Impatience	0.552	3.517	0.046	0.448

Apart from a decrease in ratings from the training to the SAME levels, mean ratings of Negative Affect increased as Relational Complexity increased. A similar pattern is observed for mean ratings of Impatience. Thus, mean data supports the single subject analysis in that there did generally appear to be a trend of less enjoyment reported in the relationally more complex levels across the majority of participants.

It must be noted how clear the foregoing effect is for all measures, across all levels, with the exception of the Training Game. It is possible that this exception is due to the fact that all participants were presented with this Training Game initially, while the order of the following three levels was randomized. Thus, no order effects are being observed for the three randomly sequenced game levels that followed training. Level 1 could be seen as a tedious game in so far as feedback was being provided on every task and no problem solving of any kind was required. Thus, we should perhaps expect to see high levels of Impatience during this game, which is relieved during other phases of the game, but which generally increases again with increasing complexity. A similar effect may explain scores of Negative Affect, although it must be borne in mind that this is not reflected in ratings of Positive Affect. Of course, the questionnaires employed are not standardized for the purposes used in the current study and this outcome reflects the difficulties typically experienced in using subjective measures. On the other hand one of the main purposes of this research is to establish precisely these types of participant reports for delivery to the gaming industry for immediate application to game development. In light of this, the questionnaire employed in the current study may prove a useful starting point for developing standardized rating scales to measure game experience and enjoyment.

A one-way repeated measures Analysis of Variance was conducted in order to examine whether the within-subjects variable of Relational Complexity, as manipulated across the four levels of the game, had any significant effect on Correct Responses recorded in the game, or on participants' reports of Positive Affect, Negative Affect, Competence or Impatience. As displayed in Table 3.6,

Relational Complexity had a significant effect on Correct Responses and ratings of Impatience, but did not have a significant effect on ratings of Positive Affect, Negative Affect or Competence. Interestingly, Positive Affect did approach a significant interaction with complexity. That is, a non significant value of $p = 0.07$ was found for Relational Complexity on ratings of Positive Affect across all levels of the game.

Experiment 3.1 aimed to; 1) model a game involving derived relations that presented different levels of difficulty to participants, and 2) to measure the effect that these different levels of difficulty had on participants' performance and their subjective ratings of enjoyment. Participants produced significantly different amounts of Correct Responses across the different games, suggesting that the different games did present varying levels of difficulty to participants. Specifically in the SAME Game, fourteen of the sixteen participants produced Correct Responses on more than 90% of all trials. In the OPPOSITE Game, twelve participants produced Correct Responses on 90% of trials, while the remaining four participants performed poorly, in two cases displaying perfect counter control by the contextual cue (i.e., achieving a score of 0). Finally, in the Mixed SAME and OPPOSITE Game only four participants produced Correct Responses on 90% of all trials, while the majority of participants scored somewhat above what would be expected by chance alone. Thus, from an examination of the number of participants achieving a high score in each individual level, it does appear that the different levels of the game presented varying levels of difficulty to participants.

Examination of the number of Correct Responses produced by participants in each game also allows for the evaluation of whether a derived relations game was successfully modeled in the current study. As most participants scored correctly on more than 90% of trials in both the SAME and OPPOSITE Games, despite the time constraints imposed, it appears that a derived relations game was successfully modeled, at least for these particular games. Specifically, participants were able to respond in accordance with derived relations between a number of arbitrary stimuli in the presence of contextual cues and time constraints, in the context of a computer game. It must be remembered also that this performance was observed in a context wherein no feedback had been given before or after the derived relations testing phase. Thus, while the game performance was not entirely novel it was entirely untrained and non-reinforced.

Most participants did not reach a 90% correct criterion in the Mixed SAME and OPPOSITE Game. Thus, it is not clear whether a derived relations game has been successfully modelled in this game. Specifically, it is difficult to know what constitutes a "pass" rate in the current game given its novelty in the general format of task presentation as well as the imposed response time constraints. It is not yet known what functional effect these features have on response accuracy, but it is at least

conceivable that it may be very difficult to generate a score of 90% on a game test format regardless of performance using other test formats. Thus, a 90% correct responding criterion on the game may not be a useful indicator of fluency as few or no subjects may ever reach that criterion. In effect, it is difficult to establish good stimulus control using the game as a test format, when there are simultaneous extraneous sources of behavioural control at work (e.g., strict time limits). In other words, it remains a possibility that the observed score rates for the Mixed SAME and OPPOSITE Game are actually quite impressive given the behaviourally challenging format of the game tasks. Interestingly, nine of the sixteen participants did produce more Correct Responses than would be expected by chance alone in the Mixed SAME and OPPOSITE Game, which suggests that participants were engaging in relational responding in a game playing environment during this game. Thus, it appears that a highly complex game involving derived relations of SAME and OPPOSITE was modelled successfully in levels 2 and 3 and likely level 4 of the current experiment.

Participants reported significantly more Impatience with the more complex games. This suggests that participants enjoyed the more complex games less than the less complex games. Additionally, while Relational Complexity did not have a significant impact on either Positive Affect, Negative Affect or Competence, there is a trend of lower Positive Affect and Competence ratings, and higher Negative Affect ratings in the later, more complex, games. Participants rated level four as both less positive and more negative than level 3 or 2. While this pattern is significant for only one of the four subjective measures, it may be argued that with larger numbers (as might be employed in a large group-design study) this pattern may have produced a significant result. Additionally, due to the fact that the questionnaire employed has not been standardized for investigating enjoyment in games, clear and significant trends are inherently difficult to identify in the data.

Despite the ANOVA results, it is important from a stimulus control and general radical behavioural perspective that we do not forget to carefully consider the single subject patterns of data obtained. A consistent, coherent trend is evident in the subjective data, which suggests that participants prefer to play the less complex games. This general trend is evident across all measures employed in the study, across all games and for most participants in each case. Indeed, previous experiments in the current thesis have also revealed, using a number of different procedures and methods of presentation, that participants consistently rate the games requiring simpler relational responding as more enjoyable. At this stage it is becoming increasingly apparent that, at least with the types of preparations developed thus far, participants do perform better at and prefer less complex games over more complex games.

3.2 Experiment 3.2

In Experiment 3.1, four game levels were presented, each one requiring participants to engage in a different type of relational responding in order to gain a high score. Importantly, it was proposed that some of those levels were more complex than others. Specifically, Level 4, which required participants to respond to randomly alternating SAME and OPPOSITE trials was conceived as the most complex level, while Level 1, which required no derived relational responding was conceived as the least complex. Level 2, the SAME level, represented a relationally simpler task than both Levels 3 and 4.

Results from Experiment 3.1 suggest that a derived relations game, involving the presentation of levels of varying complexity, was successfully modeled. However, manipulation of the complexity variable across the four levels of the game had a statistically significant effect on correct scores and ratings of Impatience, but not on any other subjective measure employed. This finding suggests that, while increasing Relational Complexity to considerably high levels does affect participants' ability to play a game successfully, it does not affect their enjoyment or experience of that game. Nevertheless, there was a trend towards lower Positive Affect and Competence ratings, and higher Negative Affect ratings in the later, more complex levels of the game, than in the less complex levels. These patterns are both coherent and consistent across levels and across participants, notwithstanding conceptual and empirical problems with the reliable quantification of subjective states.

While it now seems possible to draw the conclusion that participants prefer games that contain lower levels of Relational Complexity, it is not yet possible to draw such a conclusion about Network Latency. Indeed, a number of issues identified in previous experiments as yet remain unresolved. Specifically, the method in which Network Latency was modeled in previous experiments may have lacked ecological validity. While Latency was modeled as a fixed interval in previous studies, this does not reflect any real-world example of delay in a distributed gaming experience. Indeed, real-world Network Latency is rarely, if ever, predictable, and typically oscillates erratically during game play (Delaney, Meenaghan, Ward and Mc Loone, 2004). It has been suggested that this oscillation in Network Latency, known as 'jitter,' is much more destructive to the game playing experience than fixed delays (Delaney et al., 2004). It may be argued, therefore, that participants in the earlier experiments outlined here could have predicted the length of delays and adapted their behaviour to compensate. More specifically, in a game with programmed fixed delays participants can learn to respond appropriately across a number of delay trials. In effect, an onset of a delay (precipitated by a response) may have functioned as a discriminative stimulus for further delayed responding for a specific length of time (e.g., 2s trial with 1s delay means wait 1s before responding)

on that trial (i.e., respond again as soon as the delay passes in order to register the response). In effect, the delay trials may have represented a concurrent schedule on which rapid learning took place. The delay programmed in the previous experiments, therefore, may not have been as disruptive as if it had been of a random duration. Thus, Network Latency, as modeled in previous studies may have been perceived merely as a challenge of the game, rather than a significant nuisance. To address this issue, what is required is a more ecologically valid game in which the oscillating features can be modeled. Experiment 3.2 will involve presenting delays of varying lengths and at different points across trials. As such, these delays should be unpredictable and intrusive on the game playing experience, and thus model the effect of network jitter effectively.

In Experiment 3.1, Network Latency was held constant at zero while Relational Complexity was manipulated on a number of levels. This approach was taken in order to gain greater control over the variable of Relational Complexity. Experiment 3.2 will involve holding Relational Complexity at a constant level and manipulating latency at a number of levels. Participants will first be required to complete relational pre-training and relational training, identical to that presented in Experiment 3.1. In the game phase, participants will play one Training Game (Level1) and three games containing varying levels of Network Latency; ranging from delays on no trials (No Delay) to delays on half of all trials (Half Delay), to delays on all trials (Full Delay).

3.2.1 Method

3.2.1.1 Participants

Seventeen participants (10 female, 7 male), all first year undergraduate students aged 18 - 35, were recruited for Experiment 3.2 through personal contacts, notice board advertisements and cold calling on individuals around the University campus. Participants were not paid for their participation in the study.

3.2.1.2 Materials

All materials were identical to those used in Experiment 3.1, apart from some small changes to the Visual Basic programme that presented the experimental procedure.

3.2.1.3 Design

The experiment employed a repeated measures design with one independent variable, Network Latency, which was manipulated at three different levels. All participants were exposed to three

experimental conditions (i.e., No Delay, Half Delays, and Full Delays) in counterbalanced order. There were five dependent measures; participants' score on each level of the game, and their subjectively rated level of both positive and Negative Affect, as well as Competence and Impatience (see Kahneman, et al., 2004).

3.2.1.4 Procedure

The experiment was divided into three stages; (i) Relational Pre-training, (ii) Relational Training, and (iii) Gaming. The Relational Training and Relational Pre-training stages were identical to those used in Experiment 3.1.

(iii) Gaming

Upon successful completion of the relational training and testing procedure, participants were exposed to a computer-presented game. The game consisted of four levels. Level 1 was a training level in which participants learned how to interact with the interface. Level 2 required participants to produce derived relational performances in the presence of quasi-randomly alternating SAME and OPPOSITE contextual cues and no Network Latency. Level 3 was identical to Level 2, with the exception that simulated network delays were experienced on half of all trials presented. Level 4 involved the presentation of network delays on all trials. Importantly, the order in which these levels were presented was counterbalanced across participants. After each level a twelve-item questionnaire was presented, based on the DRM questionnaire (Kahneman, et al., 2004) which measured two constructs; Positive Affect and Negative Affect, as well as ratings for Competence and Impatience.

An identical user interface to that used in Experiment 3.1 was used in Experiment 3.2, comprising of a control panel at the bottom of the screen, with the current game level displayed in the left hand corner, the participant's score presented in the right hand corner and a button labelled 'DESTROY!' in the centre of the panel. Game characters were presented in twenty four possible locations in the area above the control panel. The procedure of an individual trial in Experiment 3.2 was identical to that in Experiment 3.1, with the exception that some trials contained delays.

The game consisted of four levels. Level 1, hereafter referred to as the Training Game, was identical to Level 1 of Experiment 3.1 and involved the presentation of the B1 and B2 stimuli from the relational training stage as game characters. Levels 2, 3 and 4 were all similar to Level 4 from Experiment 3.1, as the C1 and C2 stimuli were presented as game characters in the presence of quasi-randomly alternating SAME and OPPOSITE contextual cues.

Two functionally distinct types of simulated network delays were utilised in Experiment 3.2, in order to more accurately simulate the ‘jitter’ experienced by online game players. Type 1 delays involved delays within the response window, while Type 2 delays involved delays within the feedback window. Simulated network delays were implemented on none of the trials in Level 2 (hereafter referred to as the No Delay Game), 50% of all trials in Level 3 (hereafter referred to as the Half Delay Game), and 100% of trials in Level 4 (hereafter referred to as the Full Delay Game). Half of the delay trials in each game consisted of Type 1 delays, while the remaining trials consisted of Type 2 delays. Type 1 and Type 2 delay trials in Levels 2 and 3, as well as delayed and non-delayed trials in Level 2 were presented quasi-randomly, with no limits on how many of the same trials could be presented consecutively, apart from the maximum number of that trial type for that level.

In a trial involving a Type 1 delay; a character was presented in one of twenty four possible locations on-screen, as normal, and increased in size by 25% every 250ms in order to simulate approaching the participant. If a save or destroy response was recorded within 750ms, the programme recorded that response as normal and proceeded to the next trial. However, unlike a non-delayed trial, if no response was recorded by the time 750ms had elapsed, the character ‘froze’ on-screen. The period in which the character was ‘frozen’ constituted the delay window, and lasted for a duration 750ms. During the delay window, the character did not increase in size, the ‘Destroy’ button turned a grey colour, and both save and destroy responses were ineffectual (i.e., responses did not produce the effects observed outside the delay window). These responses were recorded unknown to participants for later analysis. Once the delay window had elapsed, (i.e., 1500ms after the trial onset), the character resumed increasing in size, and responses were enabled for the remaining 500ms of the trial.

A trial involving a Type 2 delay followed a course similar to a non-delayed trial until a response was made by the participant, at which time the character froze on-screen. The period in which the character was ‘frozen’ constituted the delay window, and lasted for a duration 500 ms. During the delay window, the ‘Destroy’ button turned a grey colour, and both save and destroy responses were ineffectual (i.e., responses did not produce the effects observed outside the delay window). These responses were recorded unknown to participants for later analysis. Once the delay window had elapsed, the character was cleared from the screen, the ‘response detected’ message was displayed and the participants score was adjusted accordingly. The following trial was then displayed.

A 12-item questionnaire was presented to participants after the completion of each individual level in the study level as a subjective measure of both positive and negative attitudes towards that level. The questionnaire used forms part of the Day Reconstruction Method (DRM; Kahnemann et

Table 3.7: Number of attempts required by each participant to pass each stage in the Relational Pre-Training and Testing procedure

Participant	Train1	Train2	Train3	Train4	Train5	Test1	Test2	Train5	Test3	Test4
1	1	1	2	1	1	1				
2	1	1	2	1	1	1				
3	1	1	3	3	4	1				
4	1	1	2	6	1	1				
5	1	1	2	1	3	1				
6	1	1	4	2	4	1				
7	1	1	2	1	1	1				
8	1	1	9	1	2	1				
9	1	1	2	1	1	1				
10	1	1	2	1	2	1				
11	1	1	2	1	2	1				
12	1	1	1	1	2	1				
13	1	1	2	1	2	1				
14	1	1	1	1	1	1				
15	1	1	2	2	6	1				
16	1	1	2	1	1	1				
17	1	1	2	1	1	1				

al., 2004).

3.2.2 Results and Discussion

3.2.2.1 Relational Pre-Training

Table 3.7 shows the number of attempts each participant took to pass each stage in the Relational Pre-Training and Testing procedure. Note that an ‘x’ in a test block represents a failure to pass the test on that attempt. Once one training block had been passed, participants advanced to the following stage and were not exposed to any further training or testing. All participants passed the first two training blocks on their first attempt. All participants then passed the following three training blocks, the majority within one or two attempts. All participants then passed the relational pre-training test on their first attempt. Thus, seventeen participants advanced to the Relational Training phase.

3.2.2.2 Relational Training

Table 3.8 shows the number of attempts each participant took to pass each stage in the Relational Training and Testing procedure. Note that an ‘x’ in a test block represents a failure to pass the test on that attempt. Once one training block had been passed, participants advanced to

Table 3.8: Number of attempts required by each participant to pass each stage in the Relational Training and Testing procedure

Participant	Train1	Train2	Train3	Test1	Train3	Test2	Train3	Test3	Train3	Test4
1	3	1	1	1						
2	1	1	1	x	1	x	2	x	1	x
3	1	3	1	x	2	x	4	x	2	x
4	1	1	1	x	1	x	1	1		
5	1	1	1	x	1	x	1	x	1	x
6	2	2	1	x	1	x	1	x	1	x
7	1	1	1	x	1	1				
8	3	1	1	x	1	1				
9	1	1	1	x	1	x	1	x	1	x
10	2	1	1	1						
11	1	1	1	1						
12	1	1	1	1						
13	1	2	2	1						
14	1	2	1	1						
15	2	1	1	x	1	x	1	1		
16	1	1	1	1						
17	1	1	1	1						

the following stage and were not exposed to any further training or testing. All Participants passed Relational Training phases one, two and three within three attempts at each phase. Eight participants passed the relational testing phase on their first exposure. Of those who did not, all passed the following training phase within two exposures. Two participants passed relational testing on their second exposure. All seven remaining participants passed the following training phase within four exposures. Two participants passed the relational test on their third exposure. All of the remaining five participants passed the following training phase within two exposures. However, none of those participants passed the final relational testing phase. Thus, a total of five participants did not pass the relational testing phase within four attempts, so their participation in the experiment was terminated. Twelve participants advanced to the gaming phase.

3.2.2.3 Gaming

Level 1 of the gaming phase was a training level in which participants learned how to interact with the user interface. All participants passed this training level and advanced on to the remaining game levels. Table 3.9 presents the total number of mouse clicks made by participants during each level of the game. It must be noted that, while the game interface appeared unresponsive for the duration of simulated network delays, the software programme did still record any mouse clicks made by participants. Table 3.9 shows that the number of clicks across a level of the game increases

Table 3.9: Total number of mouse clicks made in each level of the game by each participant

	No Delay	Half Delay	Full Delays
1	44	49	54
4	44	60	51
7	48	58	67
8	48	74	59
10	45	57	57
11	48	65	71
12	46	51	56
13	45	44	61
14	48	73	63
15	47	49	50
16	48	68	64
17	48	60	49

generally from the No Delay/baseline level to the other levels. This suggests strongly that participants are indeed responding to the delays on each trial. Participants typically produced more mouse clicks during the Half Delay and Full Delay Games than the No Delay Game. In simple terms, the presence of extra responses in the Half Delay and Full Delay Games demonstrate that the opportunity to gain extra points during the truncated response window was reinforcing.

Interestingly, there does not appear to be any trend of more or less mouse clicks made during the Half Delay than in the Full Delay Game, or in the Full Delay Game than in the Half Delay Game. Six participants produced more mouse clicks during the Full Delay Game than in the Half Delay Game, while five produced more mouse clicks in the Half Delay than in the Full Delay Game. The remaining participant produced the same amount of mouse clicks during the Half Delay and Full Delay Games. Thus, the extra delays in the Full Delay Game over the Half Delay Game do not appear to have affected participants' engagement with the game. In effect, the number of trials on which delays were administered does not appear to affect the number of responses emitted by subjects. At this stage, therefore, the effect of the delays appears to be uniform regardless of game type.

Figure 3.8 presents participants' number of Correct Responses across the No Delay, Half Delay and Full Delay levels. Six of the twelve participants produced fewer Correct Responses in the Half Delay Level than the No Delay Level. Surprisingly, four participants produced more Correct Responses in the Half Delay Level than the No Delay Level, while two participants produced an equal amount of Correct Responses in the No Delay and Half Delay Levels. Thus, while the presence of delays on half of all trials appears detrimental to the ability of half of the participants' to produce Correct Responses, other participants appear unaffected. Indeed, one third of all participants

actually achieved better scores in the presence of delays.

Four participants produced fewer Correct Responses in the Full Delay Game than the Half Delay Game, while six participants produced more Correct Responses in the Full Delay Game than the Half Delay Game. The finding that half of the participants who played the game produced more Correct Responses in the Full Delay level than the Half Delay level was not expected. It appears that these participants (P7, P8, P10, P12, P13 and P16) had less trouble making the appropriate derived relational response when there was a delay on every trial than when there was a delay on half of the trials. It is possible that these participants found the Full Delay Game more reinforcing or engaging than the Half Delay Game. This possibility is examined below in the discussion of subjective ratings.

Five participants (P1, P8, P12, P15 and P16) produced fewer Correct Responses in the Full Delay Game than the No Delay Game, while four participants produced more Correct Responses in the Full Delay Game. Thus, while there seems to be a difference in the number of Correct Responses made across the no delay and half delay levels, there appeared to be little difference in scores between the level containing no delays and the level containing the most delays.

Interestingly, there appears to be two main patterns evident in participants' scoring across levels. Specifically, the four participants who produced more Correct Responses in the Half Delay Game than the No Delay Game (P4, P14, P15, P17), were the same four participants who produced fewer Correct Responses in the Full Delay Game than the Half Delay Game. That these participants produced fewer Correct Responses in the Full Delay Game than the Half Delay Game was expected, insofar as it is in keeping with the general prediction that fewer Correct Responses should be produced when more network delays are employed. However, it was not predicted that participants would produce more Correct Responses during the Half Delay Game than in the No Delay Game. Indeed, these participants reached a 90% correct criterion when there were delays present on half of all trials. It appears that these participants had little trouble producing the correct derived relational response in the presence of network delays on half of all trials, and actually found this task less difficult than when there were no delays presented. This pattern of responding is difficult to account for. The converse pattern is evident for another group of participants (P8, P10, P12, P13, P16), who produced less Correct Responses in the half delay than the No Delay Game, but also produced more Correct Responses in the Full Delay Game than the Half Delay Game. Once again, that these participants produced fewer Correct Responses in the Half Delay Game than the No Delay Game was expected. However, it was not predicted that participants would produce more Correct Responses during the Full Delay Game than in the Half Delay Game. It appears that these participants found it more difficult to produce the correct derived relational response when network

Table 3.10: Mean Correct Responses across all across the No Delay, Half Delay and Full Delay levels (n=12)

	No Delay	Half Delay	Full Delay
Correct Responses	35.58333	33.25	35

delays were present on half of all trials than on all trials.

Overall there does not appear to be a consistent effect of Network Latency on participants' Correct Responses. It is possible that the effect of Network Latency on performance may vary across participants in tandem with levels of other variables, such as experience at on-line game playing in the presence of latency, extent of history of derived relational responding in similar contexts, or as yet unknown physiological variables. In effect, the patterns observed illustrate the high variability in participants' performance under different levels of latency presented and as such it is difficult to make recommendations or generalize from these results.

Table 3.10 presents mean Correct Responses across the No Delay, Half Delay and Full Delay levels. Mean Correct Responses are relatively consistent across all three levels, indicating that participants were equally successful on all levels of the game. These results are consistent with the single subject analysis. Indeed, there does not appear to be a consistent pattern of more or less Correct Responses in any level of the game.

Figure 3.10 presents participants' ratings of Positive Affect, Competence, Negative Affect and Impatience across the No Delay, Half Delay and Full Delay Games. One participant rated the Half Delay Game as higher in Positive Affect than the No Delay Game, while six participants rated the Half Delay Game as less positive than the No Delay Game. So, the Half Delay Game appeared to be rated as less positive in general than the No Delay Game. Three participants rated the Full Delay Game as higher in Positive Affect than the Half Delay Game, four participants rated the Full Delay Game as less positive than the Half Delay Game, while five participants rated these games as similar in terms of Positive Affect. Thus, it appears from ratings of Positive Affect that participants preferred the No Delay Game over the Half Delay Game, but there was no such distinction between the Half Delay and Full Delay Games.

Nine out of twelve participants rated the Half Delay level as higher in Negative Affect than the No Delay level. None of the twelve participants rated the Half Delay level as less negative than the No Delay level. Thus, Negative Affect ratings mirror those of Positive Affect in that most participants appeared to prefer the No Delay Game over the Half Delay Game. Two participants rated the Full Delay Game as higher in Negative Affect than the Half Delay Game, while six participants rated the Full Delay Game as lower in Negative Affect than the Half Delay Game.

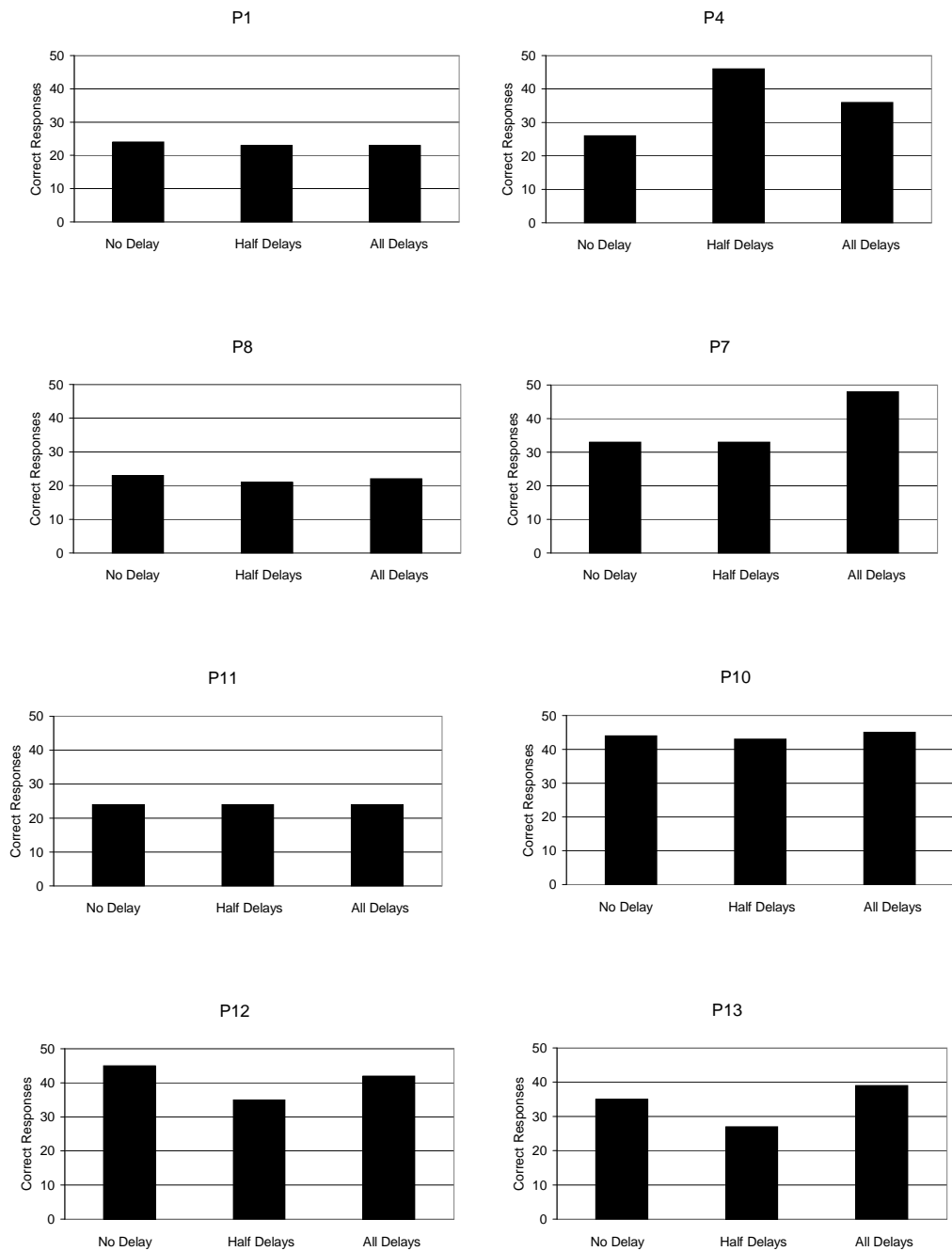


Figure 3.8: Correct Responses across the No Delay, Half Delay and Full Delay levels for all participants

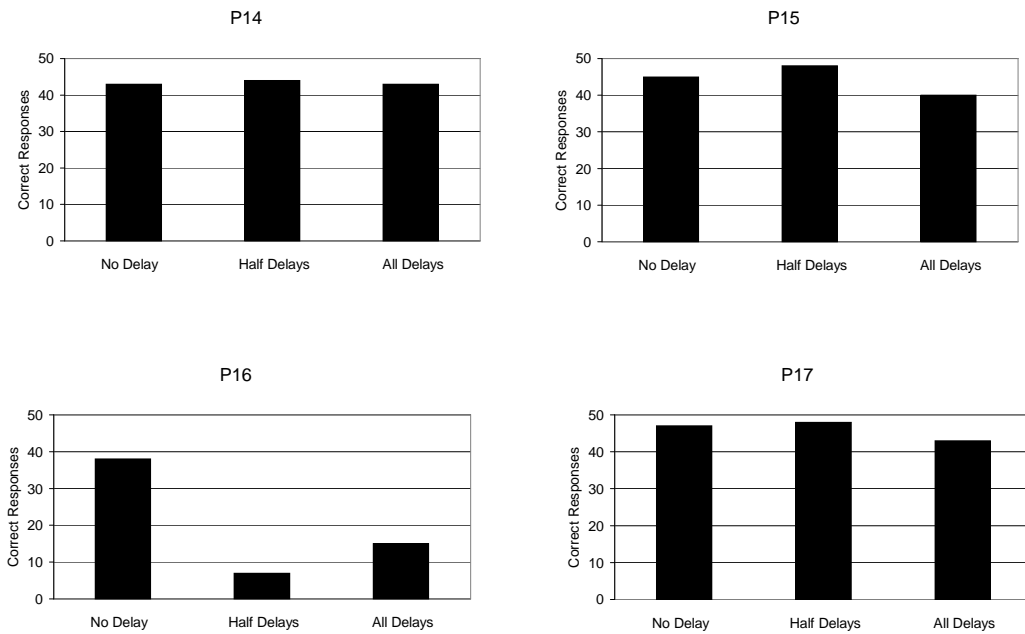


Figure 3.9: Figure 3.8 continued

Thus, while it appears that participants consistently find the presence of delays as more negative than no delays, the presence of delays on every trial was not rated as more negative than the presence of delay on only half of all trials. Indeed, unexpectedly, six of twelve participants found that delays on every trial was less negative than delays on half of the trials presented. Importantly, Negative Affect ratings appear consistent with Positive Affect ratings in that participants appear to prefer the No Delay level over the Half Delay level but did not express such a distinction between the Half Delay and Full Delay levels.

Two out of twelve participants rated the Half Delay Game as higher for Competence than the No Delay Game, while six participants rated the Half Delay Game as lower in terms of Competence. One participant rated the Full Delay Game as higher in Competence than the Half Delay Game, four participants rated the Full Delay Game as lower in terms of Competence, while seven participants produced similar ratings of Competence across both of these levels. Thus, similarly to ratings of Positive and Negative Affect, participants appear to prefer No Delay over some delay but there was not such a clear distinction in ratings between the Half Delay and Full Delay Games.

Four of the twelve participants rated the Half Delay Game as higher for Impatience than the No Delay Game, while seven participants rated the Half Delay Game similar in terms of Impatience to the No Delay Game. Four participants rated the Full Delay level as higher in terms of Impatience than the Half Delay Game. Four participants rated the Full Delay Game as lower in terms of

Impatience, while four participants produced similar ratings of Impatience across these two games. Thus, there does not appear to be a clear pattern of Impatience ratings across the three levels. However, it must be noted that Impatience ratings were very low across all levels, thus indicating a possible floor effect.

Overall, subjective ratings from the DRM questionnaire appear to suggest that participants preferred a game containing no delays over some delays, but there is not such a distinction in ratings between the Half Delay and Full Delay Games. This pattern is evident from ratings of Positive Affect, Negative Affect and Competence, while there appears to be no consistent pattern to ratings of Impatience across levels.

It was mentioned above that six participants (P7, P8, P10, P12, P13 and P16) produced more Correct Responses in the Full Delay Game than the Half Delay Game. It was suggested that these participants may have had less trouble making the appropriate derived relational response when there was a delay on every trial than when there was a delay on half of the trials. It is possible that these participants found the Full Delay Game more reinforcing or engaging than the Half Delay Game. In order to assess this point, the subjective ratings for these participants were examined (see figure 3.10). It appears that these participants did not rate the Full Delay Game as more positive and less negative than the Half Delay Game. Indeed, subjective ratings appeared consistent across these two levels for this group of participants. It must be noted that participants' self-reports using the 12-item rating scale have been erratic across all experiments in which it has been employed. Thus, it is difficult to ascertain whether this variability is due to the fact that the scale has not been standardised for the purposes of investigating enjoyment in games, or due to poor experimental control over variables.

Table 3.11 Displays mean ratings of Positive Affect, Competence, Negative Affect, and Impatience across the No Delay, Half Delay and Full Delay levels. There was a consistent drop in ratings of both Positive Affect and Competence across all three games, suggesting that participants enjoyed the games containing more Network Latency less than those which contained less Network Latency. This pattern was reflected in participants' ratings of Negative Affect and Impatience. Ratings of both Negative Affect and Impatience were greater in the Half Delay and Full Delay than in the No Delay Game. Thus, it appears as if participants enjoyed the games containing no delays more than those which did contain simulated network delays. While it does appear that the Half Delay Game was typically rated as preferable to the Full Delay Game, the difference in ratings across these two games was very small for Positive Affect, Competence and Impatience, while ratings of Negative Affect actually went in the other direction. Participants rated the Full Delay Game as

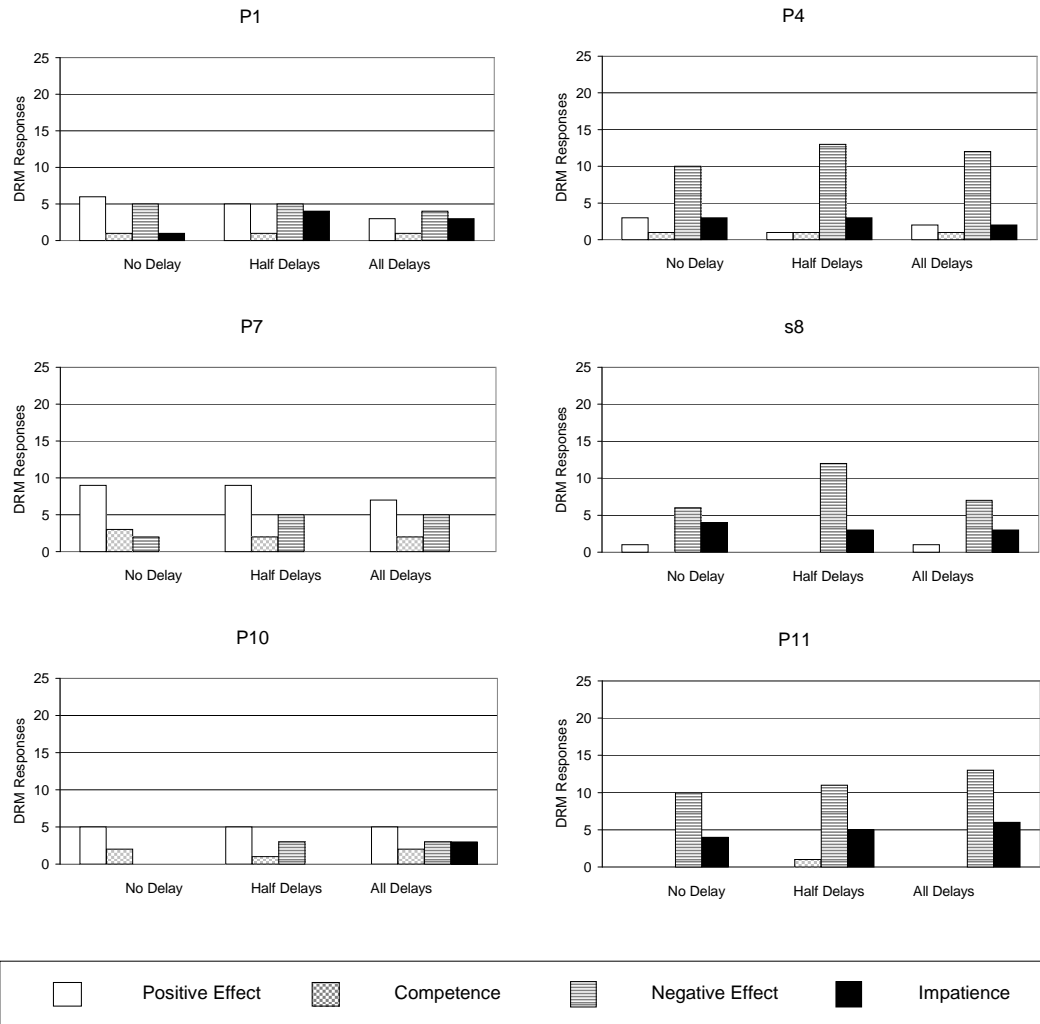


Figure 3.10: Ratings of Positive Affect, Competence, Negative Affect and Impatience across the No Delay, Half Delay and Full Delay Levels for all participants

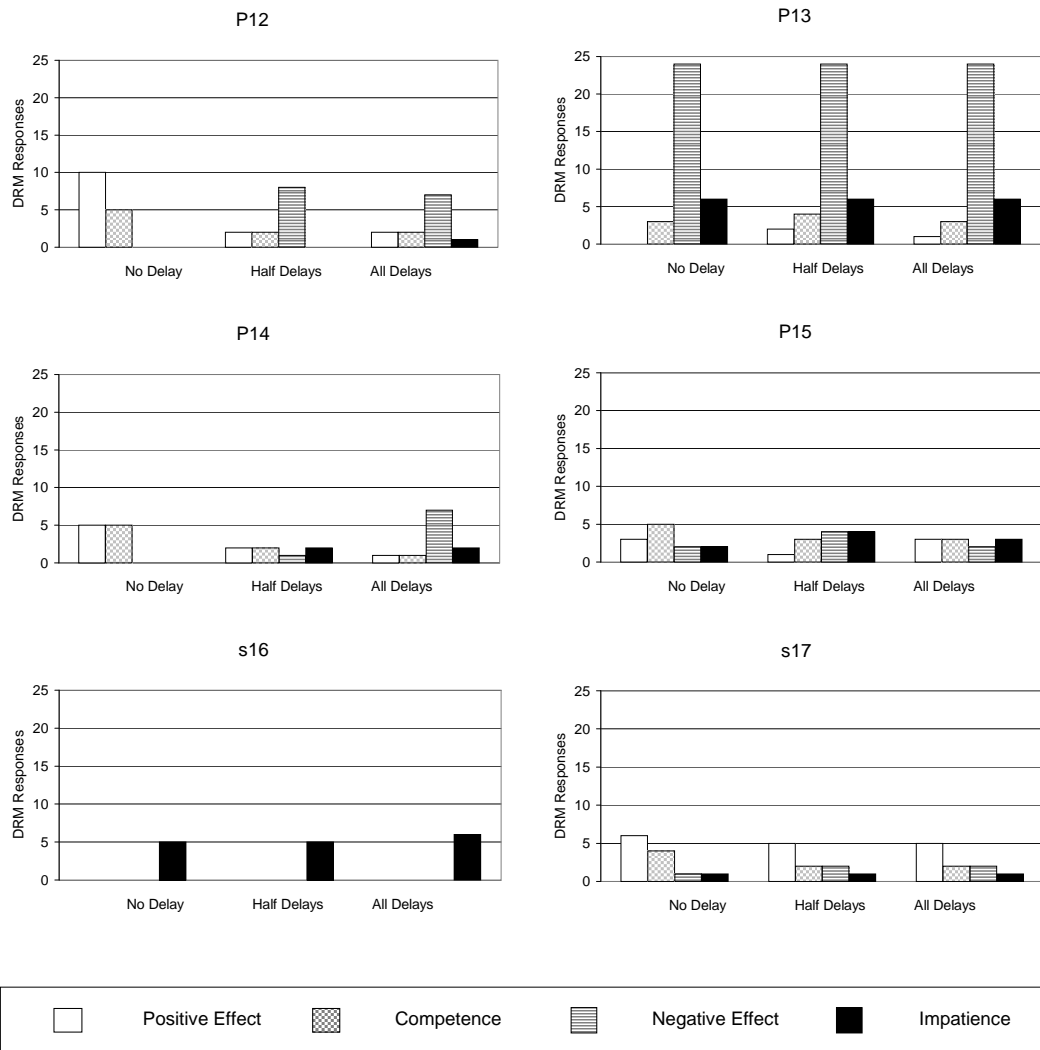


Figure 3.11: Figure 3.10 continued

Table 3.11: Mean ratings of Positive Affect, Competence, Negative Affect, and Impatience across all four levels of the game (n=12)

	No Delay	Half Delay	Full Delay
Positive Affect	4	2.666667	2.5
Competence	2.416667	1.583333	1.416667
Negative Affect	5	7.333333	7.166667
Impatience	2.166667	2.75	3

Table 3.12: Results of a one-way Repeated Measures ANOVA examining the effects of Network Latency on number of Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=12)

	Wilks' Lambda	F Value	P Value	Eta Squared
Correct Responses	0.940	0.319	0.734	0.06
Positive Affect	0.718	1.960	0.191	0.282
Competence	0.639	2.826	0.106	0.361
Negative Affect	0.468	5.681	0.022	0.532
Impatience	0.667	2.493	0.132	0.333

less negative than the Half Delay Game. Thus, it appears that while participants did prefer a game that contained no delays, the relationship between the level of delays employed and subjective ratings of enjoyment remains unclear.

A one-way repeated measures Analysis of Variance was conducted in order to assess whether the within-subjects variable of Network Latency, as manipulated across the four games, had any significant effect on Correct Responses recorded in the game, or participants' reports of Positive Affect, Negative Affect, Competence or Impatience. As displayed in Table 3.12, Network Latency had a significant effect on ratings of Negative Affect, but did not have a significant effect on Correct Responses, or ratings of Positive Affect, Competence, or Impatience. As such, it appears that the different levels of Network Latency presented Experiment 3.2 did not significantly affect participants' performance at, or enjoyment of, the games played. This finding has major implications for the computer games industry, as it suggests that a game which suffers from a large amount of delays is no less enjoyable than a game which does not suffer from any delays at all. However, as pointed out in the single subject and means analyses, there was a consistent pattern in which the Half Delay and Full Delay Games were rated as lower in Positive Affect and Competence and higher in Negative Affect and Impatience than the game which did not contain delays, across participants. While the difference in subjective ratings between the No Delay and the two delay games does appear consistent enough to produce a significant result, there was little difference in subjective ratings across the Half Delay and Full Delay levels. In effect, the similarity in ratings

across these two games may have led the ANOVA to a type 2 error.

Interestingly, the mean score on Level 2 in Experiment 3.2 (i.e., the Mixed SAME and OPPOSITE Game, with no delays) is very similar to the mean score on Level 4 in Experiment 3.1, which was also a Mixed SAME and OPPOSITE Game with no delays. This suggests that this game produces a reliable and relatively invariant game performance across a large number of subjects, and so strongly suggests that the patterns of behaviour observed here would be replicable using similar procedures and a new cohort of subjects. In addition, such a stable behaviour rate in Level 2 of Experiment 3.2 suggests that this level functioned as a sound baseline against which to examine the effect of delay across subsequent levels in the experiment.

3.3 General Discussion

The current study was conducted in order to gain greater control over the variables of Relational Complexity and Network Latency than achieved in previous studies. The two variables of interest were examined separately in Experiments 3.1 and 3.2 respectively, so that the effects of each could be identified independently.

Experiment 3.1 aimed to; 1) model a game involving derived relations that 2) presented different levels of difficulty to participants, and 3) measured the effect that these different levels of difficulty had on participants' performance and their subjective ratings of enjoyment. In Experiment 3.1, Relational Complexity was examined in terms of derived SAME and OPPOSITE relations. This represents a more complex form of derived relational responding than that employed in previous studies. Importantly, it was argued in the introduction that the use of SAME and OPPOSITE relations provides us with a functional definition of level of complexity in a game. Specifically, an OPPOSITE relation requires a more complex form of relational responding than a SAME relation (see Barnes & Hampson, 1993; O'Hora et al., 2002; Steele & Hayes, 1991; Whelan & Barnes-Holmes, 2004; Whelan et al., 2006).

Experiment 3.1 found that participants produced significantly fewer Correct Responses on the relationally more complex games than in the less complex games, suggesting that those games did represent significantly different levels of difficulty for participants. In addition, it was noted that the majority of participants produced scores in the SAME and OPPOSITE Game that would typically constitute a passing criterion on a relational testing phase. While this level of correct responding was not reached on the most complex level (i.e., the Mixed SAME and OPPOSITE Game), more than half of the participants did produce more Correct Responses in this game than would be

expected by chance. Thus, it does appear that a game involving derived relational responding was modeled successfully in Experiment 3.1.

In addition, there was a trend of higher negative and lower positive subjective ratings to the more complex levels. Indeed, this pattern was linear, in that the SAME Game was rated as more positive and less negative than the OPPOSITE Game. The OPPOSITE game, in turn, was rated as more positive and less negative than the Mixed SAME and OPPOSITE Game. While this pattern was not significant at the 0.05 level, there was a tendency for participants to prefer relationally simpler games to more complex games. Experiment 3.1 was devised in order to clarify the impact of Relational Complexity on performance and enjoyment of computer gaming, as previous studies were affected by both ceiling effects (i.e., Experiments 2.1 and 2.2) and floor effects (i.e., Experiment 2.4). Of the previous experiments, only Experiment 2.3 appears to have represented a successful manipulation of the complexity variable, insofar as scores differed significantly across levels. Results of both subjective and objective measures from that experiment parallel those from Experiment 3.1. In both experiments participants produced lower scores on the relationally more complex levels, while rating those levels as less positive and more negative than the less complex levels. Thus, it appears that Experiment 3.1 may support the previous findings. However, it must be noted that the patterns observed were not universal in either the current or previous experiments. It is possible that the effect may vary across participants in tandem with levels of other variables, such as experience at game playing, extent of history of derived relational responding in similar contexts, or as yet unknown physiological variables. Thus, individual subject variables may be leading to large variability in the trajectory of the various behavioural outcomes across phases and participants. Under these circumstances, it is difficult to make generalisations for the research literature or the gaming industry as to the general levels of complexity that will produce decrements in performance and game enjoyment for most or all participants. Nevertheless, taking into account results from the current and previous experiments, it appears that participants prefer to play relationally less complex games.

The foregoing conclusion is undoubtedly a surprising one. It suggests that Relational Complexity may not be reinforcing for many game players. However, this does not mean that all complexity is not reinforcing. Rather, it may be that complexity involving the solving of a cognitive problem, ‘on the hoof’ (as seen in the current experiments) is not reinforcing. While some individuals may choose games involving complex cognitive tasks (i.e., strategy games) over simpler games, it appears that most subjects do not respond this way in the current experiments. Further research on this topic should involve examining further dimensions of complexity such as the number

of stimuli and manipulations of response window durations. These issues are addressed in the following chapter.

Experiment 3.2 examined the impact of Network Latency on game performance and enjoyment through games that varied from containing no delays to games which contained delays on all trials. Similar to the strategy adopted for examining Relational Complexity in Experiment 3.1, Network Latency was examined independently of Relational Complexity in Experiment 3.2. In addition, 'jitter, (i.e., the unpredictable variation in the length of network delays) was modeled in a more ecologically valid method than previously achieved. Participants did not produce significantly more or less Correct Responses across the three levels in the game. In addition, Network Latency only had a significant effect on one of the four subjective measures employed. However, the single subject analysis revealed that the group design statistics masked some trends in the data, which were not evident for all participants, but which were quite consistent for a large number of participants. Specifically, it appears that while the No Delay Game was rated as more positive, higher in Competence, less negative and lower in Impatience than either the Half Delay or Full Delay Games, these levels were rated as similar to each other across all four subjective measures. These results suggest that while participants prefer to play games which do not suffer from any Network Latency, a game which suffers from a large amount of latency is not less enjoyable than a game which suffers from a modest amount of latency. This finding may have implications for the gaming industry, in that it demonstrates the necessity for games designers to eliminate the effects of latency on game-play, as a small amount of latency appears to be just as detrimental as a large amount of latency.

Results of Experiment 3.2 appear consistent with the combined results of Experiments 2.1, 2.2, 2.3 and 2.4. In Experiments 2.1 and 2.2, participants who played a game containing Network Latency achieved significantly lower scores on that game than those who played a game containing no delays. In addition, ratings of difficulty and frustration were significantly higher for participants who played the games which contained simulated network delays. In Experiments 2.3 and 2.4, participants were presented with a game containing either short 0.5s delays or long 1s delays. No significant difference was found between these groups, on any measure employed in the study. Thus, the current study demonstrates the findings of Experiments 2.1, 2.2, 2.3 and 2.4 in one single experiment, and supports the conclusion that while the presence of delay in a game is destructive to the game playing experience, increasing the length of that delay beyond 0.5s does not have any further negative effects.

Importantly, the finding that a small amount of latency appears to be just as detrimental as a

lot of latency is not consistent with findings from the field of engineering (i.e., Delaney et al., 2004; Hashimoto & Ishibashi, 2006; Vaghi et al., 1999), that suggest that incrementally increasing the level of delay in a game leads to a progressively less ‘playable’ game. However, it must be pointed out that the minimum delay employed in the current studies was 500ms. As demonstrated by Delaney et al., (2004) Hashimoto and Ishibashi, (2006) and Vaghi et al., (1999), a 500ms network delay would be considered quite large and disruptive to game-play. Engineering researchers have investigated much shorter delays, to the point where they are not perceptible by users. Thus, it may be the case that incrementally increasing delay correlates with decreasing enjoyment of a game up to a point of approximately 500ms, after which effects levels off.

It is worth reminding the reader at this point that network jitter was modeled in a more ecologically valid way in Experiment 2 than was achieved in the previous experiments presented thus far. Specifically, delays occurred at two possible points in a trial and lasted for two possible durations. Therefore, these delays should have been unpredictable and more intrusive on the game playing experience. Thus, the current manipulations of the delay variable more closely model delay and jitter in online game-play in real-world settings. However, it could be argued that a more technologically advanced method of modeling jitter, such as that employed by Delaney et al. (2004) (i.e., in which the onset and duration of network delays are truly randomized), would have provided a more complete analysis. While this may be the case, the current research is interested in examining the broad effects of latency on user experience, rather than scrutinizing the minutia of particular latency oscillations. Because network delays in Experiment 3.2 were varied sufficiently in order to be unpredictable and intrusive on the game playing experience, they were sufficient for the basic purposes of comparing games containing three different levels and distributions of delay. The current work, therefore, should be seen as a starting point in the psychological investigation of the effects of Network Latency on game playing, rather than a means of providing final conclusions on the absolute relationship between delay and game experience in a general sense. It is the responsibility of future researchers to extend this analysis to include manipulations of the whole spectrum of latency lengths and oscillations in order to more fully examine this issue.

Interestingly, the finding that participants produced a significantly different amount of Correct Responses across different levels of the game in Experiment 3.1 implies that the game format may provide a novel measure of fluency with derived relations in behavioral research more generally. Specifically, the unique game environment required participants to respond with both accuracy and speed in order to gain a high score. Indeed, such a methodology may be useful for people in applied contexts who are not just interested in acquisition of derived relations but in application of those

relational training techniques in real world contexts. For example, the game could represent a user-friendly method of assessing relational skills in a range of children with developmental delays. In such a preparation, the game could be used in place of a standard relational testing phase. Furthermore, the game format could be used as an intervention to improve the skill of deriving relations as suggested by literature on Multiple exemplar training (Berens & Hayes, 2007; Hayes, Barnes-Holmes, & Roche, 2001a; Murphy, Barnes-Holmes, & Barnes-Holmes, 2005; Gómez, López, Martín, Barnes-Holmes, & Barnes-Holmes, 2007).

The effects found in the current experiments are more varied than an experimental psychologist would hope for, insofar as order in the data is the main objective of any study into behavioural control. However, it is important to trust the data and the fact that behavioural outcomes always represent environmental conditions (i.e., the “rat is always right”). Indeed, it is scientifically conservative to respect data that are clear, but at odds with the scientist’s expectations. When variability is observed it is easy to infer all sorts of participant variables that are beyond manipulations (e.g., personality, intelligence etc). However, the objective is to gain control over the various measures while manipulating the variables of interest (i.e., delay and complexity). The variability in the various performances and outcome measures must at this stage be assumed to reflect the complexity of the variables affecting these outcomes across subjects. Thus, in principle, it should be possible to gain control over these variables. However, as a starting point it would be unwise to attempt to gain control over multiple variables simultaneously, particularly those that may themselves differ for different subjects. A more practical starting point may be to attempt to gain greater control over these variables by refining the conceptualisation of the experimental variables (i.e., complexity, enjoyment). Of course, further interactions between these variables, as defined thus far, surely occur with different game types, under various conditions. But, those interactions would appear impossible to generalize because of the infinite varieties of complexity and delay possible. Those within the industry would do well to conduct these specific analyses for their individual purposes. Future researchers engaged in developing particular games for particular audiences, however, will be armed with the crucial information about the functional relationship between delay and user performance and enjoyment, as well as informed about relational properties and keeping games relatively relationally simple.

Chapter 4

Experiments 4.1-4.3

4.1 Experiment 4.1

A number of general findings have emerged over the course of research conducted thus far. Firstly, the presence of any Network Latency in a game was found to be detrimental to the game playing experience, in that it affected both the scores obtained in, and subjective ratings of, those games. Increasing the length of delays in games generally did not have a significant effect on either Correct Responses made or subjective ratings. Secondly, it was generally the case that participants produced lower scores on the relationally more complex games and rated those games as less positive and more negative than the less complex games. However, these patterns were not observed universally across all participants and game types, and indeed conflicting patterns were sometimes observed across different experiments.

Considering the findings of the foregoing chapters, it may not be prudent to indefinitely continue pursuing the relationships between the variables of Relational Complexity and Network Latency on computer game experience and enjoyment. The variability in the various performances and outcome measures must at this stage be assumed to reflect the complexity inherent in the variables of interest. Of course, further interactions between these variables, as defined thus far, surely occur with different game types, under various conditions. But, those interactions would appear impossible to generalize because of the infinite varieties of complexity and Network Latency possible.

One important goal of the current research is to inform the research and development activities of a wide range of people involved in game production and sales. Thus, findings must not only be clear and replicable, but represented in general terms for a wide variety of scientists. The problem of

generalizing findings from the current research for a wider audience (e.g., engineers, graphics designers, etc.) are compounded by the fact that the definitions of the concepts of enjoyment, complexity and even games themselves are not technical in nature, but are defined differently by different researchers. Thus, findings from the current research are inherently difficult to generalise to specific instances of other games, whether commercially developed or part of another similar research programme. As such, the computer games industry may have some trouble gleaning knowledge from these findings, if what are sought are clear statements on the linear and stable relationships between the variables of Network Latency, complexity and gaming enjoyment. The previous chapters served to show how complex a task it would be to arrive at such conclusions. While the investigations conducted were worthwhile in demonstrating these experimental challenges, it may not be fruitful to continue to pursue the explication of these highly complex relationships between the variables of complexity and delay.

In the research that is reported in the current chapter the problem of inadequacies in definitional terms will be addressed. In particular, the research will consider alternative measures of gaming enjoyment and complexity that should make future research into the effects of game complexity and Network Latency easier to measure, predict and control. More specifically, the issue of what constitutes enjoyment in a computer gaming context is an ongoing concern for the current research. It is possible that activities that would be considered extremely unenjoyable in other contexts, may be reported as enjoyable in the context of game playing. For example, a game that causes the player a great deal of frustration may still be reported as enjoyable, if the player completes or performs well at that game. It is possible that any number of diverse factors such as achievement, progression, gaining a high score, the aesthetics of graphics, or some combination of some or all of these may contribute to the game players' enjoyment of a game. However, without a technical definition of what exactly constitutes enjoyment, it is difficult to compare games on this measure and arguably it is not even prudent to embark on a programme of investigation without these issues having been considered.

The measurement of enjoyment is especially difficult for two main reasons. Firstly, the lack of psychometric tests standardised for the purpose of measuring enjoyment in a computer game playing context. While the questionnaire employed in the foregoing research (DRM; Kahneman et al., 2004) was developed as a method of assessing emotional responses to specific recent events, and was validated with a large sample (n=909), the scale used in the current research was only a small part of a larger instrument. Indeed, there are no reliability and validity coefficients currently available for the entire instrument, never mind the specific scale employed in the current research. As such, it is

difficult to interpret the results obtained from using this scale in the current context. Secondly, verbal reports obtained from paper and pencil tests are not particularly reliable as predictors of future behaviour and assessing their validity represents a major conceptual challenge. More specifically, the results derived from the current questionnaire should have some predictive value for game enjoyment and likely correlate with verbal reports and other measures of enjoyment. However, they may not capture aspects of enjoyment which fall outside a common sense or other definition of the term. According to behaviour analysts this is because verbal reports are governed by contingencies that are established by the question asked rather than the construct (e.g., “enjoyment”) being asked about. These contingencies include things like experimental demand and desirability effects. As such, it is difficult to know in the absence of a functional analysis which features of the game are controlling the ratings.

It must be noted that verbal reports are not typically relied upon by behaviour analysts as measures of the contingencies that control reported behaviour, due to the widely reported poor relationship between verbal reports and overt behaviour (see Cabello & O’Hora, 2002; Critchfield & Epting, 1998; Dymond & Barnes, 1997; Perone, 1988, for further discussion). Specifically, what an individual says and what an individual actually does are often incongruent and when analysed do not correlate. A famous example of this comes from LaPierre (1934), who travelled across America with a Chinese couple, eating in restaurants and staying in motels. A questionnaire was later sent to each of the establishments in which they had stayed or eaten. Of the 128 replies he received from proprietors 90% reported they would not serve a Chinese couple. As a result of the incongruence between self reports and overt behaviour, behaviour analysts have typically relied upon overt behavioural measures and avoided verbal reports as experimental measures. In other words, when attempting to predict behaviour, it has proven more fruitful to observe how that individual will behave in similar circumstances, than to ask them about their intentions.

Despite the controversial status of verbal reports within the experimental analysis of behaviour, there has been a small research effort to develop reliable verbal report measures that are consistent with behavioural analytic practice. Some researchers (e.g., Perone, 1988) argue that verbal reports “may provide the only practical means of observing certain forms of behavior” (p71). However, Perone also warned that verbal reports may be used inappropriately as a short-cut around the difficulties of a true behavioural analysis. Nevertheless, verbal reports have contributed to a large amount of empirical evidence demonstrating how different types of verbal regulation such as counting, describing or planning, may be consistently related to participants’ performances (Barnes & Keenan, 1989, 1993; Holland, 1958; Leander, Lippman, & Meyer, 1968; Lowe, Harzem, & Hughes,

1978). More recently, a technique known as protocol analysis or the “talk-aloud” procedure (Ericsson & Simon, 1984; see also Hayes, 1986) has been suggested as a possible technique for analyzing verbal reports from a behaviour analytic perspective. The technique requires participants to say aloud everything they are saying to themselves privately, while responding to the particular task that is being investigated. This verbal output is then recorded, transcribed and coded. Coding involves creating a structure of categories that are relevant to both the requirements of the task and the researchers’ hypotheses about the interaction of verbal rules and direct contingencies in the context of the task being examined. These categories are then refined and scored in relation to task performance.

However, this technique has not been widely tested or explored (but see Cabello, Luciano, Gomez, & Barnes-Holmes, 2004; Cabello & O’Hora, 2002). Thus, despite these efforts, most behaviour analytic researchers appear to remain cautious about the use of verbal reports and tests that rely upon them until their value as a behaviour analytic tool has been demonstrated. In any case, protocol analysis would not appear to be suited for gaming research. That is, most computer games require a high level of response fluency under relatively challenging conditions (Rollings & Adams, 2003). Thus, a talk aloud procedure would create response competition that would interfere with both the game playing and the resulting experience. More importantly, it may make little scientific sense to pursue a verbal reporting procedure that may alert us to additional sources of behavioural control when the critical dependant measure has not been well defined. Thus, the current chapter will instead reconsider more carefully the issue of the measurement of enjoyment in more satisfactory terms.

The current research will also more closely consider the definition of complexity. In particular, there is no agreed-upon definition of what constitutes a complex game and what differentiates a more complex game from a less complex game. A huge range of variables may be appealed to in this definition, ranging from the level of derived relational responding required to interact successfully with the game characters, to factors such as speed of stimulus presentation, the number of different characters in a game, the number of possible functions that game characters can have, the ease of use of the graphic interface, to the fine motor skills required to interact with a game.

In addition, regardless of which definition of complexity is employed, the scientific psychologist, in attempting to build a model or analogue of a game, is obliged to create a “stripped down” version of the game, free of all unessential or confounding features in order to study its core processes. As a result, there will always be a certain lack of ecological validity in modelled games compared to commercially available games as used by the general public. Parenthetically, this emphasis on

experimental rigour can become a burden, because it further inhibits our ability to generalise from experimental findings to commercially available games. It should not be surprising, therefore, if the effects of complexity vary with even the slightest differences across two games employed in different experiments, or across different research labs employing entirely different games. As such, it may be difficult to generalise the findings or compare findings across research from different disciplines (e.g., psychology and software engineering).

The current research has suggested one definition in relational terms that has yielded some interesting findings. Conceiving of complexity in terms of derived relations has been useful in expanding the literature on derived relations, in that it has made a contribution to that field. Specifically, the previous experiments have illustrated the application of the derived relations concept to understanding behaviour in the real world, outside of education and special needs interventions (see Dixon, Marley, & Jacobs, 2003a; MacLin, Dixon, & Hayes, 1999, for further examples of such applications with gambling machines). As such, the work conducted to date is of conceptual and theoretical importance to researchers in that field. However, defining complexity solely in terms of derived relations has not proved to be a fruitful strategy for examining complexity in computer games. What may be needed at this point is a simpler definition of complexity with greater face validity, perhaps in terms of speed of stimulus presentation and the number of stimuli involved.

It would appear that speed of stimulus presentation and brevity of response windows appear to be a common feature of popular games, with speed typically increases as the player proceeds from level to level within the game. In addition, the “difficulty” level of a game is often defined largely by the response rate demanded to successfully interact with that level. Moreover, for many games response fluency represents a good working definition of complexity or “difficulty”. Indeed, classic arcade games such as *Pacman*, and *Space Invaders* clearly involve high response rates to achieve the highest scores. Interestingly, these games do not typically involve any obvious Relational Complexity. Thus, an examination of complexity in terms of response rate and Number of Stimuli may allow for a more basic analysis of game complexity in computer gaming than Relational Complexity (the latter may be suitable only for highly complex or specific types of games). A more basic analysis may lead to more stable relationships being identified, and thus clearer conclusions being drawn. Such a definition would also have the added advantage that a simple game with the simplest possible conceptualisation would make a good starting point for exploring the novel behavioural and psychophysiological measures discussed below. In effect, an analysis of game complexity in terms of speed and Number of Stimuli may be more likely to yield reliable relationships between complexity and enjoyment. Indeed, these relationships should be even clearer

given the alterations to the means of assessing enjoyment that will be explained later in this chapter.

The current research will also reconsider the means of assessing enjoyment in operational terms. Traditional behavioural methods, such as evaluating the reinforcing properties of different games, represent an obvious means of assessing preference and may prove to be more reliable measures of enjoyment in games. For example, simply presenting a participant with a number of games and later asking them to choose which game they wished to play again should represent a reliable method for judging which game that participant enjoyed the most. In addition, physiological measures may provide valuable real-time information that could be used for evaluating the enjoyment experienced by game players.

While the use of verbal reports within behaviour analysis is controversial, they may still provide valuable information that is difficult to measure using traditional methods. However, if they are to be used in further studies in the current research programme, it may prove beneficial to develop a method of assessing their validity. Interestingly, the novel behavioural measures of enjoyment discussed above may provide the opportunity to assess the utility of the verbal report measures used thus far. For example, an experimental preparation could involve presenting a participant with a number of different games and asking them to provide verbal ratings of each game directly after playing. Then, once all games are played, the participant could be asked to choose which game they wished to play again. This choice represents the participants' favourite game, so the overt behaviour of choosing which game they wish to play could be used to validate the verbal reports. Specifically, if the game that is rated as highest in enjoyment on the verbal reports consistently predicted which game would later be chosen, then the verbal report measures could be considered a useful tool.

Psychophysiological arousal levels may also represent an operational definition and measure of game enjoyment. Psychophysiological recording methods represent a means of measuring participants' responses without the need for verbal reports or overt behaviour on the part of the participant. Electrodermal activity (EDA) is the most widely used psychophysiological measure due to its low cost, non-invasiveness, relative ease of measurement and quantification, and more importantly, its sensitivity to psychological states and processes (Dawson et al., 2000). The use of EDA involves placing electrodes on the skin of participants' hands, in order to measure changes in electrical conductance, (changes in resistance can also be measured, but this is not common practice. See Dawson et al., 2000, for a discussion of this point) on the surface of the skin. This change in skin conductance has been linked to arousal of the autonomic nervous system (ANS). Importantly EDA, has proved valuable in a wide range of psychological research, from examining basic processes such as emotion, arousal and attention (Dawson et al., 2000), to derived relational responding (i.e., Roche

& Barnes, 1995a,b, 1997) to applied psychopathological studies (i.e., Fung, Raine, Loeber, Lynam, Steinhauer, Venables, & Stouthamer-Loeber, 2005; Schlenker, Cohen, Hubmann, Mohr, Wahlheim, Watzl, & Werther, 1995; Turpin & Clements, 1993), to studies of engagement with computer games (i.e., Mandryk & Inkpen, 2004; Prendinger, Mori, & Ishizuka, 2005; Scheirer, Fernandez, Klein, & Picard, 2002).

Eccrine sweat glands are found all over the body, with particular density on the palms of our hands and the soles of our feet. These glands are primarily used for thermoregulation, but those located on the palmar and plantar surfaces have been linked more closely with grasping behaviour. It has been suggested that these glands are responsive to significant emotional stimuli (Edelberg, 1972). EDA recording techniques involve measuring the changes in electrical signals sent to these glands by the ANS. Specifically, two electrodes are attached to the distal phalanges of the middle and index fingers of the participants' non-dominant hand, and an electrical current is passed between these two points. Changes in the conductivity of the skin are detected by the electrodes and recorded using an amplifier, polygraph and computer software.

There are two common approaches to using EDA in psychological research (Dawson et al., 2000). The first approach involves measuring Skin Conductance Responses (SCR), which are the momentary increases in conductance linked to the presentation of specific stimuli. One of the most widely used paradigms in EDA research involves measuring the initial orienting response found on EDA when a novel stimulus is presented. Typically, on repeated presentations of the same stimulus, habituation and eventually extinction are found on this response. The second approach to using EDA in psychological research involves measuring Skin Conductance Levels (SCL), which refer to the absolute level of conductance at a given moment. Such research examines SCLs over a specific period of time, such as during an ongoing task or presentation of a chronic stimulus, and compares different SCLs recorded during the completion of different tasks.

EDA has been closely linked with the psychological concepts of attention and emotion. A large number of studies have reported that tasks involving effortful processing (i.e., Zahn et al., 1999), mental workload (Verwey & Veltman, 1996) and sustained attention (Grim, 1967; Smallwood, Davies, Heim, Finnigan, Sudberry, O'Connor, & Obonsawin, 2004; O'Connell, Bellgrove, Dockree, & Robertson, 2004) produce higher SCRs and SCLs than tasks which do not involve such effortful processing. In addition, sustained attention on an ongoing task seems to lead to increased SCL over the course of that task (Bohlin, 1976; Lacey, Kagan, Lacey, & Moss, 1963; Smallwood, Davies, Heim, Finnigan, Sudberry, O'Connor, & Obonsawin, 2004; O'Connell, Bellgrove, Dockree, & Robertson, 2004) and reverses the decline in SCL over time normally associated with repeated presentation of a

stimulus (i.e., habituation). Importantly, as computer games appear to be tasks of this kind, patterns of EDA levels across time may allow the discrimination of tasks on which participants display sustained attention and those on which they do not.

In addition, a large number of studies report the ability to determine emotionality, in general, from EDA (Levenson & Gottman, 1983, 1985). However, the problem of distinguishing specific emotions from each other is more complicated. Specifically, while SCRs and SCL can be used to determine whether a participant is calm or emotionally aroused, it is difficult to distinguish between two aroused emotional states (Cahill, Ward, Marsden, & Johnson, 2001; Herbelin, Benzaki, Riquier, Renault, & Thalmann, 2004), for example frustration and enjoyment. (It must be noted that some initial progress has been made in distinguishing emotions using multiple psychophysiological measures (see Allanson & Fairclough, 2004; Collet, Vernet-Maury, Delhomme, & Dittmar, 2005; Picard, 1997; Scheirer, Fernandez, Klein, & Picard, 2002). However, these investigations are preliminary and involve complex mathematical modeling, the discussion of which is beyond the scope of the current thesis). The inability to distinguish between the EDA signals of participants displaying these very different emotions would appear to question the suitability of the measure as an index of enjoyment in a gaming context.

Fortunately, from a Behaviour Analytic perspective, it is not necessary to know in advance what the particular pattern of EDA activation for a participant who is enjoying themselves looks like. It is not necessary to identify what the “qualia” of enjoyment entail or how they might be conceptualised. What is at stake is the identification of a method to measure and/or predict the behaviours of interest; in this case the participants’ game playing and their choice of games given a choice. Recall the example, given above, of a game where participants are presented with a number of different games, and are later asked to choose which game they would like to play again. If EDA were recorded during each game, it would be possible to later examine SCLs for the different game periods and determine whether a discernable difference existed between the games that participants later chose to play again, and those they didn’t. If such a distinction were found, it could be considered that EDA served to function as a measure of the reinforcing properties of those games. Most psychologists would loosely refer to these properties as enjoyable properties.

The previous chapters investigated the effects of both Network Latency and game complexity on game playing experience and enjoyment. However, it does not seem possible to continue analysing these variables concurrently, especially since the focus of the current chapter lies with assessing a number of novel measures of game playing enjoyment. It would appear that measuring the effects of two independent variables, while simultaneously evaluating the usefulness of a number of dependent

measures, would be extremely complex and counter-productive. A simpler experimental design is necessary if the novel behavioural and psychophysiological measures are to produce any clear and comprehensive data. As such, it may prove more productive to focus on game complexity, and present participants with a number of games that differ in the complexity involved in playing those games. Complexity in the current studies is defined in two ways; 1) speed of stimulus presentation, and; 2) number of different stimuli presented within the level.

In Experiment 4.1, six games will be presented to participants. Each game will consist of a Speed of Presentation (slow, medium, or fast) and a Number of Stimuli (2, or 6). These games will be presented to participants in a randomised order. In each game, two possible response options are available: participants can either click on the stimulus itself to ‘Save’ it, or click on a button labelled “Destroy” in order to ‘Destroy’ that stimulus. After each of the six games, participants will be asked to rate that level using the 12-item scale from the DRM (Kahneman et al., 2004). SCL will be recorded from participants for the entire duration of the experiment, with stimulus markers sent from the computer game programme to the polygraph at the beginning and end of each level, in order to aid data analysis.

4.1.1 Method

4.1.1.1 Participants

Twelve participants, (10 male, 2 female) all aged 18-25 were recruited from a sample of convenience for the study.

4.1.1.2 Materials

Experiment 4.1 was conducted in two rooms, specially designed for the purposes of physiological recording. The room in which the participant was located was shielded from external static electrical fields and electromagnetic radiation using a Faraday cage. All electrical devices, apart from the LCD computer screen (resolution 1024 x 768) used to present stimuli, were located in an adjoining room. Stimuli were presented on the computer screen via Microsoft Visual Basic 6.0 software (c.f., Dixon & MacLin, 2003; Cabello, Barnes-Holmes, O’Hora, & Stewart, 2002) and a PC, which also recorded the nature and timing of stimulus presentation and participant responses. Standard computer headphones were used to present auditory stimuli and feedback to participants. EDA was recorded using Silver Silver chloride (AgAgCl) electrodes (diameter 5mm) and the EDA module of BrainVision© Amplifier and BrainVision© data acquisition software, using another PC. Stimulus

markers were sent from the Visual Basic programme to the polygraph using the PC's printer port output. Data was later analysed using BrainVision© Analyser software and a standard spreadsheet package.

4.1.1.3 Design

Experiment 4.1 employed a 2x3 factorial repeated measures design. The independent variables were the number of different stimuli presented in a game (2-stimuli or 6-stimuli), and the speed at which stimuli were presented in a game (slow, medium, or fast). Thus, there were six game types in total (i.e., 2-stimuli slow, 2-stimuli medium, 2-stimuli fast, 6-stimuli slow, 6-stimuli medium, 6-stimuli fast) All participants were exposed to all six games in a counterbalanced order. There were four dependent measures employed; participants' electrodermal activity (EDA), as measured by Skin Conductance Level (SCL), participants' total number of Correct Responses on each game, and their subjectively rated level of both Positive and Negative Affect (see Kahneman et al., 2004).

4.1.1.4 Procedure

Setting up and Instructions

Twelve participants, all aged 18-35 were recruited for the study. All participants were brought to a quiet, windowless room that was specially designed and built for the purpose of physiological recording. This room was constructed using a Faraday cage in order to shield physiological recording devices from external static electrical fields and electromagnetic radiation. The room itself contained only an LCD computer screen (resolution 1024 x 768), a mouse, keyboard, set of headphones and the BrainVision © Amplifier and electrodes. The computers used for presenting stimuli on the LCD screen, and recording all data, were located in an adjacent room.

Participants were first seated in front of the computer monitor and given a brief overview as to the nature of the study. Instructions were given that only the mouse was to be used for interacting with the computer programme, and that this should be done with the participants' dominant hand. Silver Silver chloride (AgAgCl) electrodes (diameter 5mm) were then connected to the distal phalanges of the index and middle fingers of participants' non-dominant hand. A skin conductance electrode gel was applied and the electrodes were fixed in place with tape. At this point the experimenter verified whether the electrodes had been attached properly and whether the software was recording. The headphones were then placed over the participants' ears and the volume on the computer was tested to ensure that participants could hear the presented stimuli comfortably.

Participants were then directed to read the on-screen instructions and begin the experiment.

A number of instructions were then presented on-screen, one at a time, in a large font. Once each screen had been read, participants could proceed to the next one by pressing the continue button. The instructions presented were as follows, where the text within quotation marks represents the text presented on each individual screen:

“Thank you for agreeing to participate in this research. First, MAKE SURE YOUR PHONE IS SWITCHED OFF !!!!!!!”

“During the course of the experiment it is important that you relax and do not move your arm or hand that is attached to the polygraph recorder”

“Please take a few moments to get yourself comfortable, making sure your arm is relaxed, then click continue.”

At this point a 45 second window was presented, in which participants could not interact with the computer. This was intended both to allow participants to relax, so that their EDA readings would “stabilise” in the absence of auditory or visual stimulation, and to obtain a baseline reading of that participants’ arousal. The following instructions were presented on-screen during this window:

“The polygraph is currently calibrating a baseline reading. Please wait, keeping as still as possible.”

Once this 45 second window had elapsed, participants were prompted to click a button on-screen labeled, ‘continue,’ in order to proceed. A number of instructions were then presented, which detailed the mechanics of how to score points in the game. These instructions were presented on-screen, one at a time, in a large font. Once each screen had been read, participants could proceed to the next one by pressing the continue button. The instructions presented were as follows:

“PLEASE REMEMBER to keep your arm that is attached to the polygraph AS STILL AS POSSIBLE for the duration of the experiment”

“In this game you will have to gain points by saving or destroying the objects which approach you”

“You can SAVE the objects by clicking on them with your mouse pointer”

“You can DESTROY the objects by clicking on the destroy button below”

At this point a graphic representation of the control bar used in the game was displayed, with the DESTROY button pointed out. Participants could proceed to the next instructions by pressing the continue button:

“Please note your score is located in the bottom right hand corner of the screen”

At this point a graphic representation of the control bar used in the game was displayed, with the Score Display pointed out. Participants were then prompted to click the on-screen, ‘continue,’ button in order to begin the game.

Game

Six games were presented in a quasi-randomized order that was counter-balanced across all participants. All games employed a simple user interface consisting of a control bar at the bottom of the screen and a large blank white space in which game characters were presented (see figure 2.2). The control bar consisted of; a level indicator located on the left hand side of the bar, a score indicator on the right hand side of the bar, and a button labeled, “DESTROY!!!” displayed centrally on the control bar. The level indicator specified how many games had been played up to that point. For example, the level indicator always displayed ‘Level 1’ in the first game that was presented, regardless of which game was actually presented.

At the beginning of each game, participant’s were prompted to click on a button labeled, ‘continue,’ in order begin that game. The screen was then cleared and a message was displayed in a large font size stating, “Ready???????” This message remained on-screen for 2s, before being replaced by a message stating, “Go,” which was displayed for 1s. Once this time had elapsed, participants were presented with the 48 trials specific to that game.

Every trial, across all games, followed the same basic structure. A stimulus, or game character, was presented in a quasi-randomised position on-screen. This game character was then increased in size by 10% a total of eight times, in order to simulate approaching the screen. The amount of time between each increase in size was dependent on the game being played. The amount of time the character remained on-screen after the last increase in size was exactly the same as the time between increases in size. Thus, the length of time given for a response to each trial was dependent on how quickly the character appeared to approach the screen, which was dependent on the game being played. For example, in the fast game stimuli increased in size every 100ms, so each trial lasted a maximum of 800ms. If the participant made a response within the 800ms, the trial was terminated and the programme proceeded to the inter-trial interval.

Importantly, the characters presented were arbitrary stimuli; simple coloured geometric shapes, (i.e., a green triangle, a red rectangle, a blue square) and novel sets of stimuli were presented in each level. Thus, before playing a level, participants were unaware of which stimuli needed to be saved and which needed to be destroyed in order to earn points. Only through trial-by-trial learning

within a game level could the participant consistently produce the experimenter-defined correct response to a particular trial type.

On the presentation of a game character, participants were required to make a stimulus discrimination response. Each character could be either *Saved*, by clicking on that character with the mouse pointer, or *Destroyed*, by clicking on the, “Destroy,” button. If a response was defined as correct, the participant’s score was increased by one point, and a positive sound was played through the headphones. If a response was defined as incorrect, the participant’s score remained unchanged, a negative sound (an explosion) was played through the headphones, and the background colour was set to red for 100ms, before returning to its original colour. If no response was made within the required time for that game, this was defined as an incorrect response; thus, the participant’s score remained unchanged, a negative sound was played through the headphones, and the background colour was set to red for 100ms, before returning to its original colour (white). Regardless of whether the response was correct or incorrect, each trial was followed by a 200ms inter-trial interval, after which the following trial was presented.

Once 48 trials had been completed in a game, the screen was cleared and the participants’ score for that game was displayed in a large font, centrally, for a duration of 5s. When those five seconds had elapsed, participants were presented with the post-game questionnaire, which consisted of a scale from the DRM (Kahneman, et al., 2004), designed to assess subjective experience of recent events. The questionnaire was presented in the center of the screen, one question at a time. Responses were given on a six-point sliding scale. Participants were not able to move on to the following question until the current question had been on-screen for a minimum of 10 seconds. This time constraint allowed participants to consider the question, and also, importantly, functioned as a further baseline period that allowed EDA readings to stabilize between games. Questionnaires were presented after every game, and took at least 120s to complete. When all questions in the DRM had been answered, the experiment moved on to the next game, until all six games had been presented, at which point the experiment was terminated.

The main variables manipulated in the experiment were *Number of Stimuli* and *Speed of Presentation*. Number of Stimuli was manipulated at two levels (2-stimuli and 6-stimuli) and Speed was manipulated at three levels (Slow, Medium, and Fast). Thus, the six games may be described as a) 2-stimuli- slow, b) 2-stimuli-medium, c) 2-stimuli-fast, d) 6-stimuli-slow, e) 6-stimuli-medium, and f) 6-stimuli-fast.

Number of Stimuli was manipulated at two levels; 2-stimuli and 6-stimuli. In a 2-stimuli game, each trial involved the quasi-random presentation of one of two possible stimuli in a randomized

position on-screen. Thus, 24 trials of each stimulus type were presented in a 2-stimuli game. One of these stimuli was designated as requiring a *Save* response in order to earn points, while the other required a *Destroy* response. In a 6-stimuli game, each trial involved the presentation of one of six possible stimuli. Thus, eight trials of each stimulus type were presented in a 6-stimuli game. Three of these stimuli were designated as requiring a *Save* response in order to earn points, while the other three required a *Destroy* response. Importantly, the 6-stimuli games were conceptualized as more difficult than the 2-stimuli levels, irrespective of speed.

Speed was manipulated at three levels; slow, medium and fast. In a slow game, stimuli remained on-screen for 225ms before increasing in size by 10%. The amount of time the character remained on-screen after the last increase in size was the same as the interval between increases in size. Thus, in a slow game, each individual trial lasted for 1800ms. If participants did not respond within this 1800ms, the trial ended and the next trial was presented. In a medium game, stimuli remained on-screen for 150ms before increasing in size by 10%. As such, trials lasted for 1200ms in a medium game. Trials automatically ended after this 1200ms had passed if participants had not produced a response by that point. In a fast game, stimuli remained on-screen for 100ms before increasing in size by 10%. Thus, fast game trials lasted a total of 800ms. Importantly, faster trials were conceived of as more difficult than slower trials, regardless of the Number of Stimuli presented.

EDA

Skin Conductance Levels were recorded continuously from the beginning of the experiment until after it had been completed. In order to identify the segments of interest for later analysis, stimulus markers were sent to the polygraph to signal the onset and conclusion of each individual game. This was carried out by triggering the PC's digital printer port output, as facilitated by Microsoft Visual Basic 6.0 software.

There are two important reasons for using Skin Conductance Levels rather than Skin Conductance Responses in the context of analysing computer game playing. Firstly, in order to accurately measure SCRs, at least 3-4 seconds must be allowed between stimulus presentations in order that the response to a stimulus is not contaminated by the fact that the response to the previous stimulus has not yet recovered (Levinson, Edelberg, & Bridger, 1984). Indeed, SCR paradigms typically involve 20-60 second inter-trial intervals (Dawson et al., 2000). Such a preparation is not feasible when studying computer games, which are typically fast paced, with trials that last non-experimentally specified amounts of time. A computer game which involved the presentation of stimuli to the participant only every five seconds would not resemble a modern

computer game either graphically or functionally. Secondly, SCL provides a continuous record of electrodermal activity across a block of trials, much like the cumulative records used in traditional behaviour analytic studies (i.e., Skinner, 1959). Such a moment to moment analysis may reveal patterns of arousal change that may serve as signatures of future behaviour, or a guide to classifying the types of contingencies in operation in the same way in which a scalloped response schedule may be indicative of a Fixed Interval schedule of reinforcement. In other words, analysing SCLs may uncover signature patterns of arousal during game play that indicate a likelihood of that game being played again under free operant conditions. Such patterns could possibly also correlate with verbal reports.

4.1.2 Results and Discussion

Twelve participants (10 male, 2 female) completed the entire experimental procedure. Upon initial ‘eyeballing’ of the behavioural data it was apparent that two of those participants (both male) could not be included in the analysis. Both had chosen only one response option (i.e., the *Save* response) for every trial over the course of the experiment. In effect, these subjects did not engage with any of the games. It later emerged that these subjects had not read the instructions completely and were not fully aware of the different response options available during the games. Thus, scores achieved on every game were exactly the same (i.e., chance levels of producing a correct response: 24 correct, 24 incorrect). Put simply, these participants were not making a stimulus discrimination response on each trial, but merely a topographical response. Thus, data from these participants would not appear to be informative for the purposes of the current study.

Table 4.1 and Figures 4.1-4.4 present mean data for Correct Responses, Positive Affect, Negative Affect and EDA across all participants for the six games employed in the current experiment. Correct Responses refers to the total number of Correct Responses participants made to the stimuli presented in a particular level. The 12 item DRM questionnaire, which was presented to participants after each game, allowed the experimenter to calculate Negative Affect and Positive Affect scores for each game level. Participants rated whether they agreed or disagreed with a number of statements on a six point sliding scale. Six of the questions referred to negative experiences, so the answers to these six questions were summed to form the statistic Negative Affect (i.e., maximum total score = 36). Three of the remaining questions referred to positive experiences and the answers to these three questions were summed to form the statistic Positive Affect (i.e., maximum total score = 18).

EDA data was recorded continuously at a rate of 500 samples per second from the beginning of the experiment until it had been completed. Stimulus markers were sent to the polygraph to signal

Table 4.1: Mean Correct Responses, Positive Affect, Negative Affect and SCL, across all conditions in the experiment (n=10)

	Slow	Medium	Fast
Mean Correct Responses			
2-Stimuli	46.1	42.9	23.6
6-Stimuli	33.5	27.6	12.1
Mean Positive Affect			
2-Stimuli	12.3	11.7	7.8
6-Stimuli	8.5	7.6	4.4
Mean Negative Affect			
2-Stimuli	4.4	5.8	14.2
6-Stimuli	11.2	11.9	18.5
Mean SCL ($\mu\text{S}/\text{cm}^2$)			
2-Stimuli	49.06724	47.95258	49.19948
6-Stimuli	45.77253	46.63737	47.72253

the onset and conclusion of each individual level of the game. Using these stimulus markers, the segments of raw EDA data that corresponded in time to the game levels were then extracted and exported to a spreadsheet for analysis, using the BrainVision © Analyser software. As the area of the electrodes used in the current study were 0.19634954cm^2 , all EDA data first had to be transformed so that results could be presented in units of microsiemens (μS)/ cm^2 . This was carried out by multiplying all raw EDA data by 5.092958 (i.e., $1/0.19634954$).

Table 4.1 presents mean values for Correct Responses, Positive Affect, Negative Affect and Skin Conductance Levels, for all games employed in the current study. For ease of analysis these data points are represented below in a visual form for each of the main variables.

Figure 4.1 illustrates that the 2-stimuli slow game produced the highest mean number of Correct Responses, while the 6-stimuli fast game produced the lowest. Mean Correct Responses are generally higher in the games involving the presentation of 2-stimuli than the 6-stimuli games, which suggests that participants were less successful when playing a game involving a larger amount of stimuli. However, this was not true in all cases. Indeed, the 2-stimuli fast game produced a lower mean number of Correct Responses than both the 6-stimuli,slow and 6-stimuli medium games. There is a linear pattern of lower mean Correct Responses in the faster levels. In both the 2-stimuli and 6-stimuli games there are lower mean Correct Responses in the medium speed game than the slow game, and lower mean Correct Responses again in the fast game over the medium speed game. Participants appear to be less successful at the faster games.

Figure 4.2 illustrates that the 2-stimuli slow game produced the highest mean Positive Affect rating, while the 6-stimuli fast game produced the lowest mean Positive Affect rating. Mean Positive

Mean Correct Responses

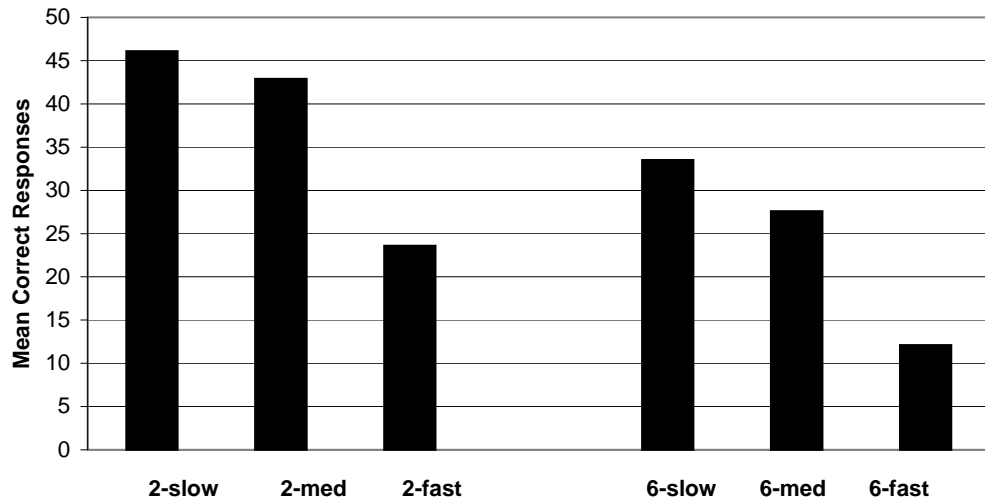


Figure 4.1: Mean Correct Responses across the six games(n=10)

Mean Positive Affect

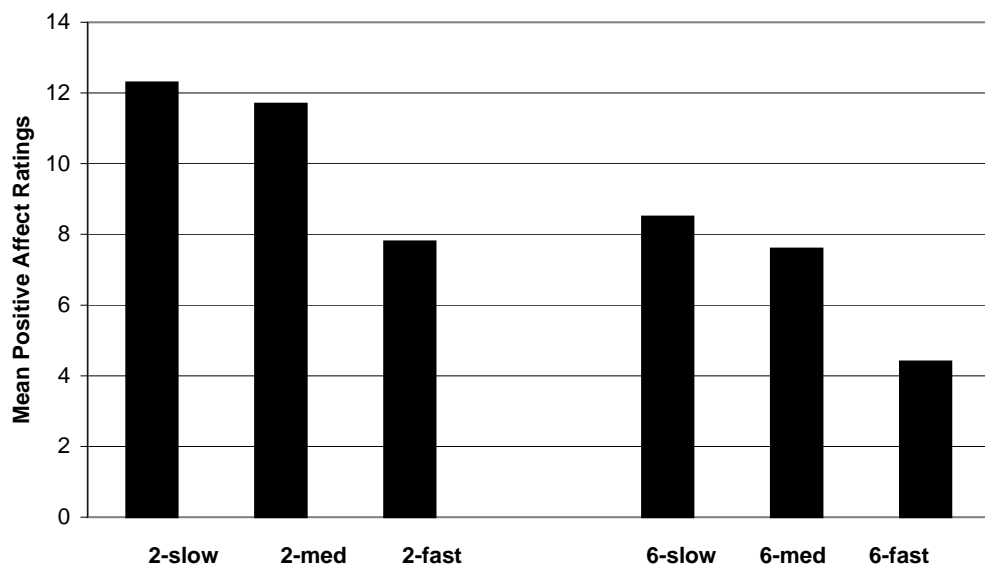


Figure 4.2: Mean Positive Affect across the six experimental conditions(n=10)

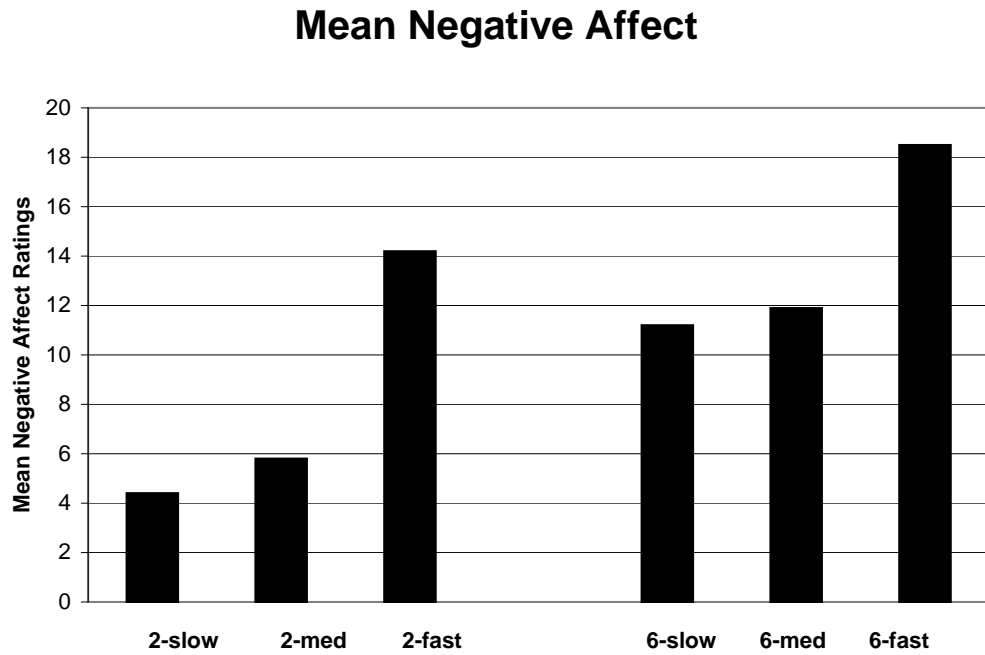


Figure 4.3: Mean Negative Affect across the six experimental conditions (n=10)

Affect ratings are generally higher in the games involving the presentation of 2-stimuli, suggesting that participants preferred the 2-stimuli games. However, this was not true in all cases. The 2-stimuli fast game produced lower ratings of Positive Affect than the 6-stimuli slow game. There is a linear relationship of lower mean Positive Affect ratings to the faster games. In both the 2-stimuli and 6-stimuli games there are lower mean Positive Affect ratings in the medium speed game than the slow game, and lower mean Positive Affect ratings again in the fast game than the medium speed game.

Figure 4.3 illustrates that the 6-stimuli fast game produced the highest mean Negative Affect rating, while the 2-stimuli slow game produced the lowest mean Negative Affect rating. Mean Negative Affect ratings are generally higher in the games involving the presentation of 6-stimuli, suggesting that participants preferred the 2-stimuli games. However, this was not true in all cases. The 2-stimuli fast game produced higher ratings of Negative Affect than both the 6-stimuli slow and 6-stimuli medium games. There is a linear relationship of higher mean Negative Affect ratings to the faster games. In both the 2-stimuli and 6-stimuli games there are higher mean Negative Affect ratings in the medium speed game than the slow game, and higher mean Negative Affect ratings again in the fast game over the medium speed game.

Importantly, the Positive Affect and Negative Affect graphs appear to mirror each other.

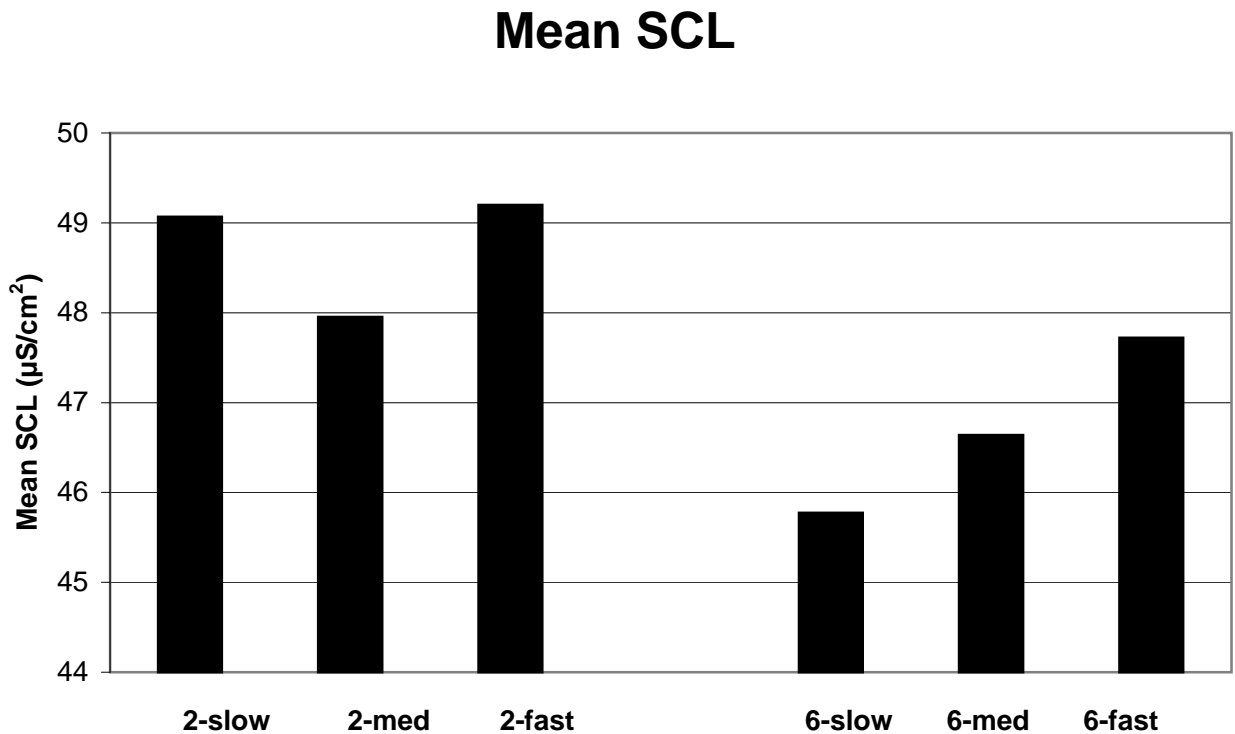


Figure 4.4: Mean SCLs across the six experimental conditions(n=10)

Games that were rated high in Negative Affect appear to have been rated low on Positive Affect. Stepping back further, it appears that the games which were rated lower on Positive Affect, and higher on Negative Affect also produced the lowest number of Correct Responses. Indeed, the pattern observed across conditions for ratings of Positive Affect is almost identical to that for Correct Responses. It appears that participants consistently rated games in which they produced a large number of incorrect responses as being less enjoyable than those on which they produced mainly Correct Responses.

Overall, participants appear to produce higher Correct Responses, higher Positive Affect ratings and lower Negative Affect ratings in the slower games. In addition, participants appear to produce higher Correct Responses, higher Positive Affect ratings and lower Negative Affect ratings in the 2-stimuli games over the 6-stimuli games.

Figure 4.4 illustrates that the 2-stimuli fast game produced the highest mean SCL readings, while the 6-stimuli slow game produced the lowest mean SCL readings. SCLs are typically higher in the 2-stimuli games than the 6-stimuli games. Indeed, all three games that involved the presentation of 2 stimuli show higher mean Skin Conductance Levels than any 6 stimuli game. There does not appear to be a consistent pattern across speeds. Within the 2-stimuli games, mean SCL dropped off

from the slow to medium game, then increases from the medium speed game to the fast game. Within the 6-stimuli games, mean SCL increases linearly as the game speed increases, across all three games.

It may be useful to compare patterns of EDA with patterns observed in the other measures employed in the study. Examining the variable of stimulus number first, there appears to be generally higher arousal levels in the 2-stimuli games than the 6-stimuli games. This correlates with a general pattern of higher Correct Responses, higher ratings of Positive Affect and lower ratings of Negative Affect in the 2-stimuli games, suggesting that the higher mean arousal levels correlate with enjoyment.

No such clear pattern of data is evident when the variable of *Speed* is examined. Within the 6-stimuli levels, mean SCLs increase from the slow to medium game, and likewise from the medium to fast game, as Correct Responses and ratings of enjoyment decrease. Thus, in the context of speed, it appears that increasing arousal is correlated with decreasing enjoyment. Further, within the 2-stimuli games, mean Correct Responses and mean Positive Affect ratings decrease linearly from slow to medium and medium to fast levels, and Negative Affect increases across those levels. This pattern is not reflected in SCL readings, where mean SCL decreases from the slow to medium levels, but then increases from the medium to fast levels. There does not appear to be any clear correlation between mean arousal levels and enjoyment in the current experiment.

It must be pointed out that even if there were a clear correlation between mean arousal levels and game enjoyment, this information would be of limited use for future research. A finding that the game which produces the highest mean SCL is the most enjoyable game, would only be useful in a situation such as Experiment 1, where a number of similar games are being compared to each other. It does not provide any valuable absolute information about any individual game examined in isolation. For example, there is no particular absolute 'level' of arousal that indicates enjoyment, as SCLs vary greatly across people due to extraneous individual differences (Dawson et al., 2000, p.208). All that would be known is that among a number of similar games, the game in which mean SCLs were highest is the most enjoyable game. Additionally, the comparative analysis across games can only be carried out after all games have been played, which limits the applicability of the technique. Regardless, the game that produced the highest mean arousal levels in Experiment 1 was not reported as the most enjoyable game, and no clear and reliable trend of this kind is obvious in the data.

While the foregoing analysis does not seem to have revealed any characteristic patterns of arousal associated with enjoyable games, this may be as a result of the assumptions of a traditional

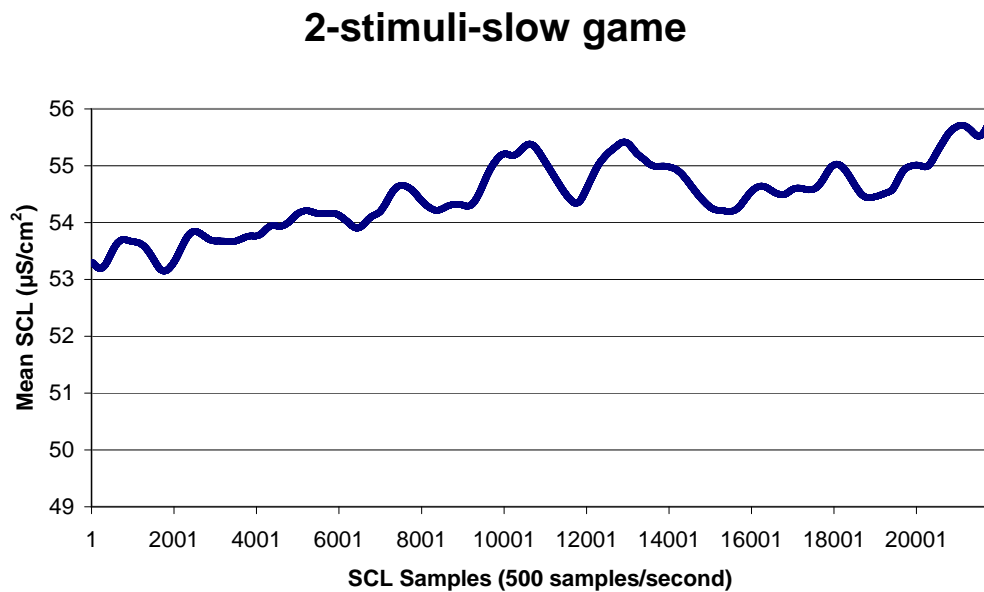


Figure 4.5: Mean SCL for all participants for the duration of the 2-stimuli-slow game(n=10)

group-design. More specifically, group analyses typically look for linear effects of variables averaged across subjects and do not examine functional differences in response rates, durations or intensities across time within individual subjects. This single subject based methodology may prove useful in the analysis of the current SCL data. Such an analysis may lead to the possibility of discovering a prototypical trajectory, or pattern of change across time, in SCLs, that correlate with enjoyable games. Thus, the remainder of the section will focus on analyzing the pattern of change in SCLs across the course of different game sessions.

A line graph was created for each participant from the raw SCL data (in microsiemens/cm²) recorded during the playing of each game. These line graphs are presented in Appendix 2 for perusal and analysis by the reader. As each participant played six games, and there were ten participants in total, a verbal description of each graph of the sixty graphs produced would render any useful overview of trends and general findings difficult to ascertain. In order to be able to arrive at clear conclusions about the effects of the main variables of interest, it was necessary to develop a method of analyzing these graphs meaningfully. As such, six line graphs were created, one for each level. Mean SCL data was calculated for each individual SCL data point across participants and a line was created from this new data. It must be noted that, as each participant spent a different amount of time playing each game, some participants' SCL data output for a particular level was longer than others. In order to solve this problem, data from all participants was cut to the length of the shortest sample for that level. Figures 4.5-4.10 present mean SCL data across all participants for each game.

2-stimuli-medium game

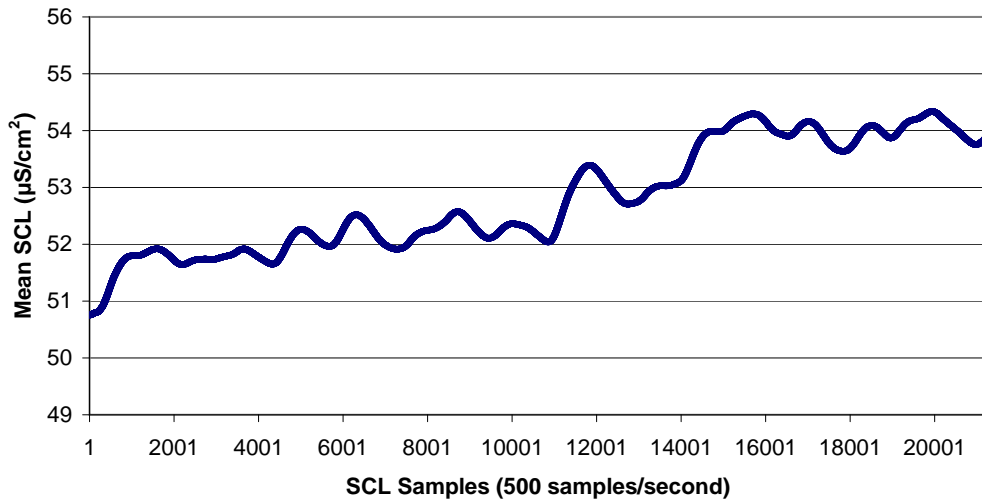


Figure 4.6: Mean SCL for all participants for the duration of the 2-stimuli-medium game(n=10)

Figure 4.5 illustrates the mean skin conductance level observed across all participants while playing the 2-stimuli-slow game. Mean SCL is relatively high and increases gradually across the course of the game. The mean score on the game was 46.1/48, indicating that the majority of participants responded appropriately to almost all trials in this game. Seven participants scored higher on this game than any other game, while two participants achieved their joint highest score on this game. Seven participants rated it as either the highest or joint highest game in terms of Positive Affect, while seven participants rated it as either the lowest or joint lowest game in terms of Negative Affect. Thus, it appears that the majority of participants produced high scores during this game, and rated it as enjoyable.

Figure 4.6 illustrates the mean skin conductance level observed across all participants while playing the 2-stimuli-medium game. Mean SCLs begin relatively low in comparison to the 2-stimuli-slow game, and show a similar gradual increase over the duration of the task. The mean score on the game was 42.9/48, indicating that the majority of participants responded appropriately to a large majority of the trials presented in this game. One participant scored higher on this game than any other game, while two participants achieved their joint highest score on this game. Four participants rated it as either the highest or joint highest game in terms of Positive Affect, while five participants rated it as either the lowest or joint lowest game in terms of Negative Affect. Thus, it appears that the majority of participants produced high scores during this game, and rated it as enjoyable.

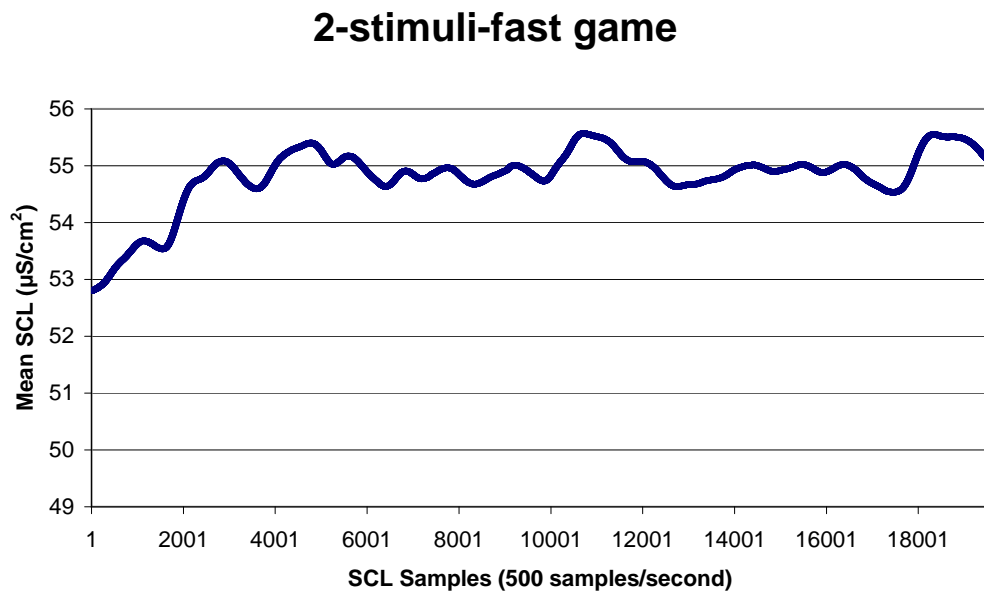


Figure 4.7: Mean SCL for all participants for the duration of the 2-stimuli-fast game($n=10$)

Figure 4.7 illustrates the mean skin conductance level observed across all participants while playing the 2-stimuli-fast game. Mean SCLs are relatively high over the course of this game. There appears to be a rapid increase in Mean SCL over the first few seconds of the game, followed by an apparent recovery period. Mean SCLs then remain relatively consistent across the remainder of the game. The mean score on the game was 23.6/48, indicating that the majority of participants had difficulty responding appropriately to the trials presented in this game. Indeed, this level of correct responding falls within that which would be expected by chance alone, due to the fact that there were only two response options for every trial. None of the ten participants rated the 2-stimuli-fast game as either the highest or joint highest game in terms of Positive Affect, while one participant rated it as either the lowest or joint lowest game in terms of Negative Affect. Thus, it appears that the majority of participants produced low scores on this game and did not rate it as enjoyable.

Figure 4.8 illustrates the mean skin conductance level observed across all participants while playing the 6-stimuli-slow game. SCLs are relatively low over the course of this game. There appears to be a rapid increase in Mean SCL over the first few seconds of the game, followed by a period where Mean SCLs remain relatively consistent. There is a further period where Mean SCLs increase towards the end of this segment of data. The mean score on the game was 33.5/48, indicating that, while participants scored above chance levels while playing this game, the majority of participants had some difficulty responding appropriately to the trials presented. Two of the ten participants rated the 6-stimuli-slow game as either the highest or joint highest game in terms of Positive Affect,

6-stimuli-slow game

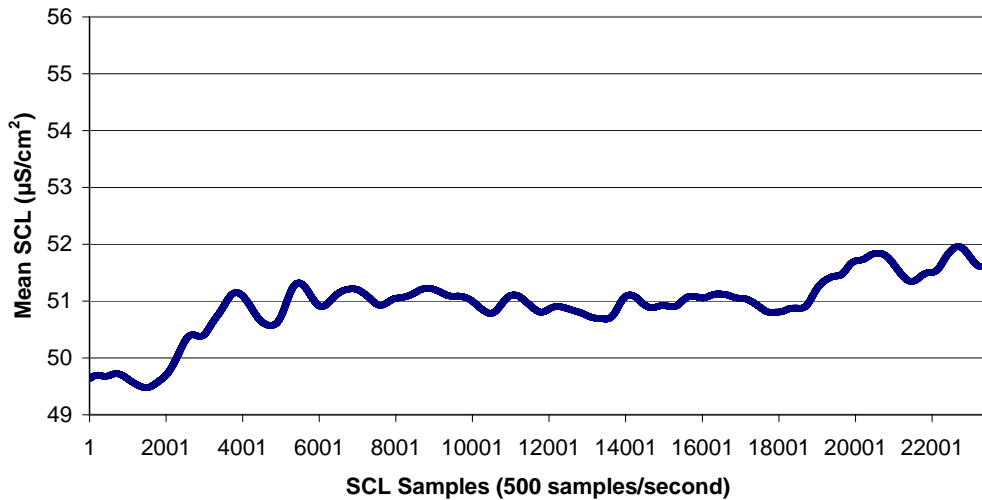


Figure 4.8: Mean SCL for all participants for the duration of the 6-stimuli-slow game(n=10)

while two participants rated it as either the lowest or joint lowest game in terms of Negative Affect. Thus, it appears that the majority of participants had some difficulty producing Correct Responses on this game, but a number of participants rated it as enjoyable.

Figure 4.9 illustrates the mean skin conductance level observed across all participants while playing the 6-stimuli-medium game. There appears to be a rapid increase in SCL over the first few seconds of the game, followed by a gradual decrease, where SCLs return to baseline. Mean SCLs then remain relatively consistent over the remainder of the game. The mean score on the game was 27.6/48, indicating that the majority of participants had difficulty responding appropriately to the trials presented in this game. Indeed, this level of correct responding falls within that which would be expected by chance alone, due to the fact that there were only two response options for every trial. One of the ten participants rated the 6-stimuli-medium game as either the highest or joint highest game in terms of Positive Affect, while one participant also rated it as either the lowest or joint lowest game in terms of Negative Affect. Thus, it appears that the majority of participants had some difficulty producing Correct Responses on this game, and did not rate it as enjoyable.

Figure 4.10 illustrates the mean skin conductance level observed across all participants while playing the 6-stimuli-fast game. There appears to be a rapid increase in SCL over the first few seconds of the game, which gradually returns to baseline over the course of the game. The mean score on the game was 12.1/48, indicating that the majority of participants had difficulty responding appropriately to the trials presented in this game. Indeed, this level of correct responding is below

6-stimuli-medium game

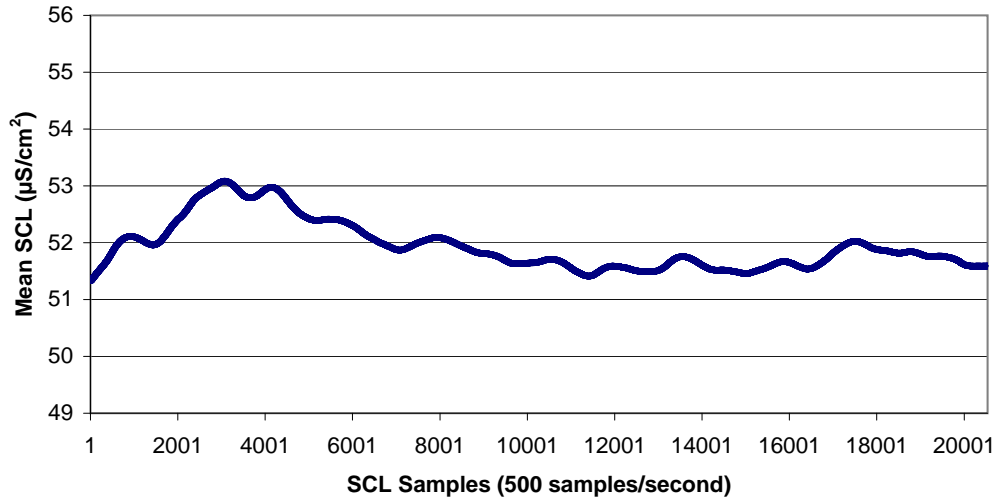


Figure 4.9: Mean SCL for all participants for the duration of the 6-stimuli-medium game(n=10)

6-stimuli-fast game

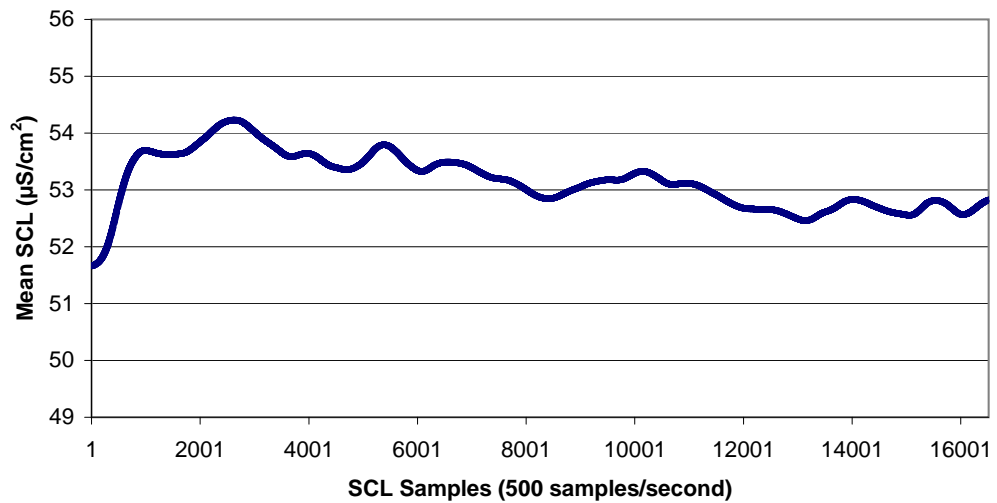


Figure 4.10: Mean SCL for all participants for the duration of the 6-stimuli-fast game(n=10)

that which would be expected by chance alone. It appears that participants struggled to register any response to the trials presented in this game. None of the ten participants rated the 6-stimuli-fast game as either the highest or joint highest game in terms of Positive Affect, while no participant rated it as either the lowest or joint lowest game in terms of Negative Affect. Thus, it appears that the majority of participants had great difficulty producing Correct Responses on this game, and did not rate it as enjoyable.

Overall, there appear to be three patterns evident in the data. Firstly, in both the 2-stimuli slow and medium games, SCLs increase gradually across the course of the game session, a high number of Correct Responses are achieved by participants, and these games are rated as enjoyable. Secondly, in the 2-stimuli fast game and the 6-stimuli slow game, an initial rapid increase is evident in SCLs, before some recovery, followed by a period where SCL remains relatively consistent. Participants achieved moderate to high scores on these games and a minority of players rated them as enjoyable. Finally, in the 6-stimuli medium and fast games, an initial rapid rise in SCL is observed, before a gradual decrease across the remainder of the game. Participants achieved poor to moderate scores in these games and they were not rated as enjoyable. It appears that the gradual increase in SCL observed in 2-stimuli-slow & 2-stimuli-medium games may be indicative of enjoyment in a game, while the gradual decrease in SCL observed in the 6-medium and 6-fast games may be indicative of games that were not enjoyed.

SCL graphs for each level were presented and discussed individually above. Specifically, the absolute values of the SCLs were of little concern in the foregoing analysis and presenting multiple graphs against common axes may have served to distract from the analysis of changes in arousal across time. However, it may also be interesting to calculate the abstracted slope (i.e., trends) of each graph in order to more closely examine the idea that the increase in SCL across time serves as a guide to the level of enjoyment experienced during a game. While this technique for representing the data eliminates the important shape of the SCL arousal curve, it has the advantage of providing a clear mathematical descriptor of the trend in arousal change across in SCL for each level of the game.

For the purposes of analysis, the first 5s of all SCL recordings were eliminated. Spikes in arousal during the first 5s are evident in the majority of raw data from all games and across all conditions. These spikes in arousal are typical upon the presentation of any stimulus (Dawson et al, 2000) and in the current study appear to be linked to participant's orienting to the game onset. Researchers have found that responses to any novel stimuli have typically peaked and are in decline within a maximum of 5s. Given that the current experiment was concerned with tonic SCL, (i.e., Skin Conductance Levels) rather than phasic SCL, the initial phasic 5s orienting response was eliminated

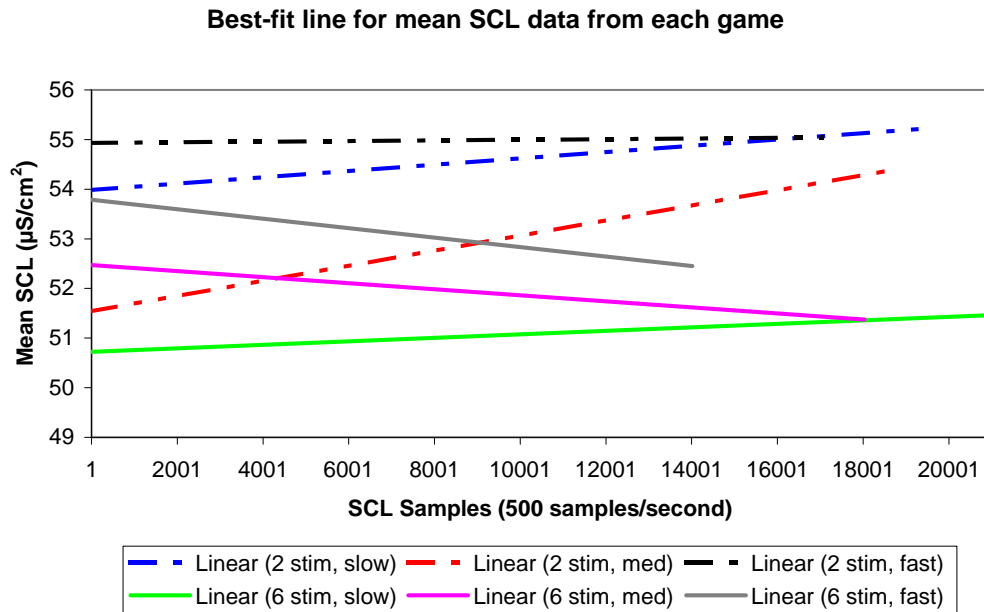


Figure 4.11: Best-fit line for mean SCL recordings from each game (n=10)

Table 4.2: Slope and correlation co-efficient values for the best-fit line of mean SCL for each game (n=10)

	Slope ($\mu\text{S}/\text{cm}^2$)/S	r^2 value
2-stimuli-slow	0.03	0.4689
2-stimuli-medium	0.1	0.8618
2-stimuli-fast	0.0035	0.0152
6-stimuli-slow	0.02	0.4026
6-stimuli-medium	-0.03	0.5156
6-stimuli-fast	-0.05	0.8162

from all data before an analysis of slopes was carried out. Results of this analysis are presented graphically in Figure 4.11.

A best-fit line was calculated for the mean SCL data from each level. These are presented in figure 4.11 in order to illustrate the average increase or decrease in arousal across the course of each game. Three separate patterns are evident. Firstly, the lines representing data from the 2-stimuli-slow, 2-stimuli-medium, and 6-stimuli-slow games display an increase in SCL across time. Secondly, the line representing the 2-stimuli-fast game appears to remain stable for the duration of the experiment. Finally, the lines representing the 6-stimuli-medium and 6-stimuli-fast games demonstrate a decrease in SCL across time. Table 4.2 displays the slope and r^2 values for each line.

Table 4.2 presents the slope and correlation co-efficient values for the best-fit line of mean SCL for each game. The lines representing four of the games (2-stimuli-slow, 2-stimuli-medium,

Table 4.3: Mean Correct Responses, Negative Affect and Positive Affect across the Slow Medium and Fast Games, with data from 2-stimuli and 6-stimuli games combined (n=10)

	Slow	Medium	Fast
Correct Responses	39.8	35.25	17.85
Positive Affect	7.8	8.85	16.35
Negative Affect	10.4	9.65	6.1

2-stimuli-fast and 6-stimuli-slow) demonstrate positive slopes. For each of these games, SCL generally increased across time. However, the slope and r^2 values for the 2-stimuli-fast game are very small. Indeed, the rise in SCL across time for this game was negligible. Two of the games (6-stimuli-medium and 6-stimuli-fast) demonstrate negative slopes. In these games, SCL generally decreased across time.

Importantly, the games in which increased in SCL across time were observed (i.e., 2-stimuli-slow, 2-stimuli-medium, and 6-stimuli-slow) were also the games most likely to be rated as highest for Positive Affect and lowest for Negative Affect. They were also the levels for which the highest mean Positive Affect ratings and the lowest mean Negative Affect ratings were observed. The games in which decreases (i.e., 6-stimuli-medium and 6-stimuli-fast games) or negligible increases (i.e., 2-stimuli-fast) in SCL across time were observed, were also the games least likely to be rated as highest for Positive Affect and lowest for Negative Affect. They were also the levels for which the lowest mean Positive Affect ratings and the highest mean Negative Affect ratings were observed. Thus, it appears, at least in the current preparation, that when an increase in SCL across time during a game is observed, that game will later be reported as enjoyable. Conversely, it appears that when a decrease in SCL across time during a game is observed, that game will not be reported as enjoyable.

In order to evaluate the individual effects of each of the two main variables in the current study, it may prove valuable to examine these independent of each other. In order to examine the main effect of Speed of Presentation, data from the 2-stimuli and 6-stimuli games are combined, so that the effects of Speed of Presentation alone, regardless of stimulus number, can be examined. For example, when examining measures of the slow game below, these represent the 2-stimuli slow game and the 6-stimuli slow game combined. In order to examine the main effect of stimulus number, data from the slow, medium and fast games are combined for both the 2-stimuli and 6-stimuli variables.

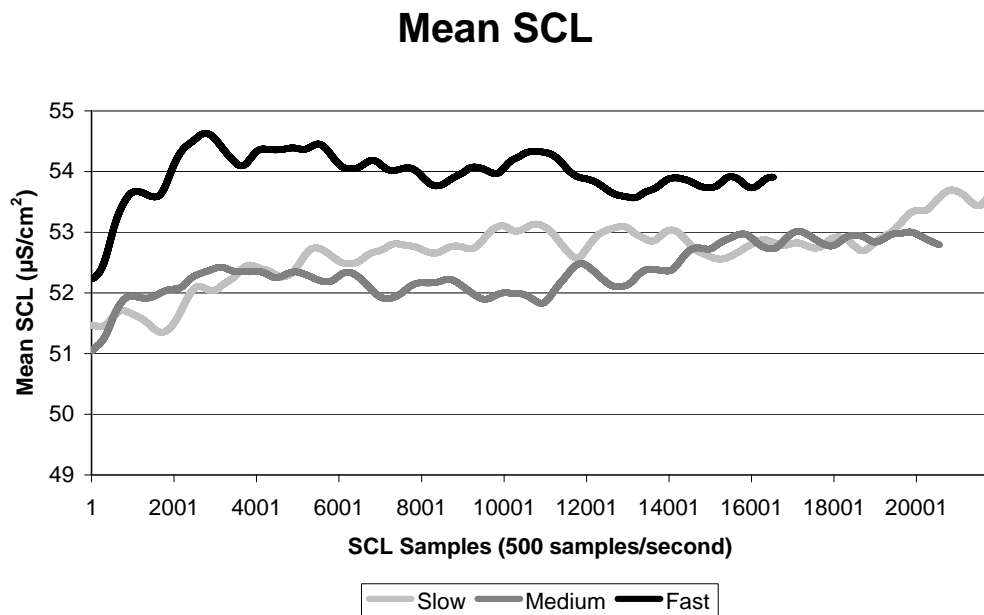


Figure 4.12: Mean SCL data across all participants, across the Slow, Medium and Fast games. Data from the 2-stimuli and 6-stimuli games are combined to produce this graph (n=10)

4.1.2.1 Main effect: Speed of Presentation

Data from the 2-stimuli and 6-stimuli games are combined and displayed in table 4.3 in order to illustrate the main effect of Speed on all four dependent measures in the current study. Correct Responses and ratings of Positive Affect decrease from the slow to medium games and from the medium games to the fast games. Ratings of Negative Affect increase across those same games. This suggests that the faster games are less enjoyable and produce fewer Correct Responses.

Figure 4.12 illustrates the main effect of Speed of Presentation on Skin Conductance Levels. Data from the 2-stimuli and 6-stimuli groups are combined to produce each line on this graph. Mean SCLs appear to rise gradually across the course of the slow games. As seen from the mean table above, the highest scores are typically achieved in the slow games and these games are rated as higher in Positive Affect and lower in Negative Affect than the faster games. Mean SCLs show an initial rapid increase in the medium games, followed by a period of recovery, which is then followed by a period of gradual increase. As illustrated by the table of means above, participants gain relatively high scores on the medium games and rate them as marginally lower in Positive Affect and higher in Negative Affect than the slow games. Mean SCLs show an initial rapid increase in the fast games, which is then followed by a gradual decrease across the course of the games. As illustrated in the table of means above, participants also produced a small number of Correct Responses in the

Table 4.4: Mean Correct Responses, Negative Affect and Positive Affect across the 2-stimuli and 6-stimuli games, with data from the Slow, Medium and Fast games combined (n=10)

	2-stimuli	6-stimuli
Correct Responses	37.53333	24.4
Negative Affect	8.133333	13.86667
Positive Affect	10.6	6.833333

Mean SCL

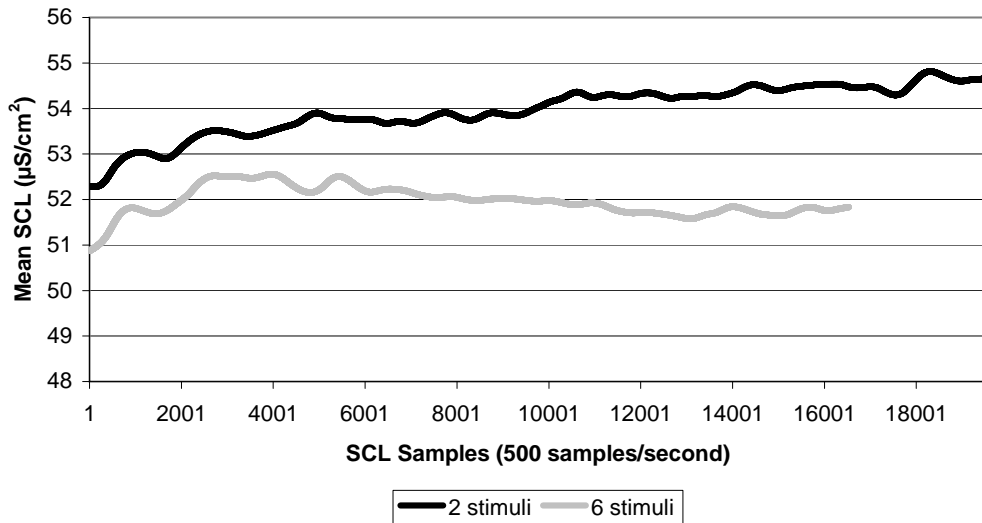


Figure 4.13: Mean data across all participants across the 2-stimuli and 6-stimuli games. Data from the slow medium and fast games are combined to produce this graph (n=10)

fast games and rated these games as low in Positive Affect and high in Negative Affect. As observed when examining levels individually, it appears there may be a link between any gradual increase in SCL ratings across time and enjoyment of that game, while any decrease in SCL ratings across time appears to correlate with participants reporting not enjoying the game.

4.1.2.2 Main Effect: Stimulus Number

Data from the slow, medium and fast games are combined and displayed in table 4.4 in order to illustrate the main effect of Stimulus Number on all four dependent measures in the current study. Correct Responses and ratings of Positive Affect are higher in the two stimuli game than the 6-stimuli game, while ratings of Negative Affect are lower in the 2-stimuli game. This suggests that participants both produced more Correct Responses in the levels involving the presentation of 2-stimuli, and preferred those levels.

Table 4.5: Results of ANOVA examining the effects of Speed of Presentation and Stimulus Number on number of Correct Responses, Positive Affect, Competence, Negative Affect and Impatience (n=10)

	Wilks' Lambda	F Value	P Value	Eta Squared
Correct Responses				
Speed	0.115	65.496	0.000*	0.885
Stimulus Number	N/A	53.86	0.000*	0.75
Interaction	0.863	1.354	0.285	0.137
Negative Affect				
Speed	0.400	12.737	0.000*	0.600
Stimulus Number	N/A	3.531	0.077	0.164
Interaction	0.970	0.265	0.77	0.030
Positive Affect				
Speed	0.471	9.542	0.002*	0.529
Stimulus Number	N/A	7.828	0.012*	0.303
Interaction	0.992	0.064	0.938	0.008

Figure 4.13 illustrates the main effect of stimulus number on Skin Conductance Levels. Data from the 2-stimuli and 6-stimuli groups are combined to produce each line on this graph. Mean SCLs for the 2-stimuli games increase gradually over the course of those games. Participants also gained higher scores in these games than the 6-stimuli games, and rated them as higher in Positive Affect and lower in Negative Affect, as shown in table 4.4 Mean SCLs for the 6-stimuli games show an initial rapid increase, followed by a gradual decrease over the course of those games. Participants also gained lower scores in those games than the two stimuli games and rated them as lower in Positive Affect and higher in Negative Affect, as shown in table 4.4. As observed when examining levels individually, it appears there may be a link between any gradual increase in SCL ratings across time and enjoyment of that game, while any decrease in SCL ratings across time appears to correlate with participants reporting not enjoying the game.

An Analysis of Variance was conducted in order to discover whether the within-subjects variables of Speed of Presentation, as manipulated across the slow, medium and fast games; and stimulus number, as manipulated across the 2-stimuli and 6-stimuli games, had any significant effect on Correct Responses recorded in the game, participants' or reports of Positive Affect or Negative Affect. The results of this analysis are presented in table 4.5.

Table 4.5 presents results of ANOVA examining the effects of Speed of Presentation and Stimulus Number on number of Correct Responses, Positive Affect, Competence, Negative Affect and Impatience. Note that where a P value is marked with an asterisk, this represents a statistically significant result. Speed of Presentation significantly affected both the number of Correct Responses made in a given game, and participants' ratings of positive and Negative Affect. Specifically,

participants scored significantly better in the slower games and rated these more positively and less negatively than the faster games. Stimulus number significantly affected the number of Correct Responses made during a game, and also participants' ratings of Positive Affect, while the effect on ratings of Negative Affect is approaching significance. Participants scored significantly better in the 2-stimuli games than the 6-stimuli games, and rated the 2-stimuli games as more positive. There was no significant interaction between the variables of stimulus number and Speed of Presentation on any of the measures employed. This implies that regardless of stimulus number, Speed of Presentation always had an effect on enjoyment and correct scores, and *vice versa*.

Previous studies found that relationally less complex games were rated as more enjoyable and produced higher Correct Responses than more complex games. In the current experiment, complexity was not defined in terms of derived relational responding, but speed of stimulus presentation, and the Number of Stimuli presented in each level. The current experiment found that participants gained significantly higher scores on simpler games (i.e., 2 stimuli rather than 6 stimuli, slow rather than fast), and rated these games as significantly higher in Positive Affect, and lower in Negative Affect than more difficult games. Thus, the current study appears to support previous findings that participants show preference for simpler games.

In addition, the current study found that there appear to be different patterns of change in Skin Conductance Levels across time for games that are enjoyed than games that are not enjoyed. Specifically, in the current study, the games in which mean SCL increased over the course of that game were later rated as higher in Positive Affect and lower in Negative Affect than the other games presented. Conversely, the games in which mean SCL decreased over the course of that game were later rated as lower in Positive Affect and higher in Negative Affect than the other games presented. Thus, it may be possible to discriminate whether a player is enjoying a particular game from an ongoing analysis of that game players' Skin Conductance Levels.

Interestingly, in the current study it appeared that enjoyment and score were closely related. That is, games on which participants produced a large amount of Correct Responses were also the games which were rated highest for Positive Affect and lowest for Negative Affect. However, separating out the effects of scoring from the effects of difficulty level represents a technical challenge. Specifically, any attempt to change the difficulty level presented in a game, in order to decrease score differences across games, would necessarily result in a different set of games. One possible way to do so would be to manipulate the scoring system employed in the games, while keeping the games the same. The reader is reminded that in some of the studies presented in previous chapters, no connection was found between game score and enjoyment. However, these

studies were complicated by the presence of a further variable (Network Latency), which may have interacted with enjoyment and score in an unforeseen manner. Experiment 1 in the current study did find a link between score and enjoyment, in the absence of complicating factors, thus it remains a possibility that score is at least a partial determinant of game enjoyment.

It is surprising to find that the easiest games presented in Experiment 4.1 lead to the highest ratings of Positive Affect, given the body of literature that links ‘challenge’ with game enjoyment (i.e., Gingold, 2005; Malone, 1982; Morlock, Yando, & Nigolean, 1985; Vorderer, Hartmann, & Klimmt, 2003). It appears that challenge is not being sought by participants in the current study, or at least that the easiest game appears to present sufficient challenge. However, the games presented in the current study were very quick (less than 2 minutes) and were only presented once each to participants. It would be interesting to know what would happen if subjects were given the opportunity to play these games repeatedly. As participants play a less challenging game repeatedly, we should expect to see changes in the ratings. Under these circumstances we may see high scores diverge from enjoyment ratings to a greater extent. Additionally, allowing game players to choose which games they wish to play at certain stages of an experiment would also provide a convenient operant measure of enjoyment that may be more reliable than verbal reports. Indeed, we may also observe changes in this operant measure across exposures.

4.2 Experiment 4.2

Experiment 4.1 demonstrated that it may be possible to discriminate whether a player is enjoying a particular game from an ongoing analysis of that game players’ Skin Conductance Levels. In addition, it was found that participants gained significantly higher scores on simpler games (i.e., 2 stimuli rather than 6 stimuli, slow rather than fast), and rated these games as significantly higher in Positive Affect, and lower in Negative Affect than more difficult games. This finding has implications for the computer games industry, as it suggests that game players prefer simple games in which they can easily gain high scores. In addition, this finding does not sit well with some of the literature on computer games, which often identifies ‘challenge’ as a key determinant in game players’ enjoyment of a game (i.e., Gingold, 2005; Malone, 1982; Morlock et al., 1985; Vorderer et al., 2003). However, two issues have been identified that may have affected these results; the fact that each game was only played once by each participant, and the uncertainty over the reliability and validity of the verbal report measure employed. Before outlining Experiment 4.2 of the current chapter, the foregoing issues require brief consideration.

All experiments conducted thus far in the current research programme involved presenting participants with a number of games that varied in terms of some particular variable of interest, and measuring the effect of this manipulation on participants' subsequent game playing behaviour and their subjective verbal reports. Typically, when complexity was examined, participants reported preference for the less complex games. However, the paradigm employed may have biased participants towards favouring the simpler games. Specifically, in every experiment conducted thus far, each game was presented once only. As participants did not have experience with any of the games, and the simpler games afforded greater opportunity for participants to produce fluent and correct responding on their first exposure to that game, this may have led to a preference for those simpler games. It could be argued that, if participants had sufficient experience with the more complex games to produce the same level of correct responding observed in the simpler games, preferences may shift towards those more complex games. As such, it may prove valuable to present all games repeatedly to participants, in order to observe whether preferences shift away from the simpler games, towards the more complex games across exposures. Under these circumstances we may see high scores diverge from enjoyment ratings to a greater extent. Such a preparation would also have the advantage of increased ecological validity, as game players rarely play each level in a commercial game only once. In effect, preference for computer games usually develops across multiple exposures to the game under real-world conditions.

While Experiment 4.1 identified a possible psychophysiological method of discriminating whether a player is enjoying a particular game, the validity of this measure is somewhat reliant on its correlation with subjective verbal report measures of enjoyment in games. The finding that participants appear to prefer simple games over complex games is also based on verbal reports. However, as mentioned in the introduction to this chapter, verbal reports are typically not relied upon in behaviour analytic research as reliable guides to future behavior or private experience. An operant measure of preference would provide us with a more robust method of evaluating participants' enjoyment of games that vary in complexity, and would also allow, in subsequent experiments, for further evaluation of the utility of EDA as a measure of enjoyment in games. An example of such a procedure could involve presenting participants with a number of games in a randomized order, before presenting the opportunity for participants to play any one of these games again in a free-operant preparation. The choice made by participants can be considered, by definition, to indicate their preferred game. Repeating this procedure a number of times would allow for an analysis of any shift in preference over time, as discussed above.

Experiment 4.2 will involve presenting the same six games to participants as those used in

Experiment 4.1. However, the procedure will involve an adaptation and extension of that employed in Experiment 4.1. Experiment 4.2 will consist of four stages. Participants will first be presented with all 6 games in a randomised order, as before. However, participants will then be given a choice of which game they wish to play next. Once participants have played their chosen game, they must then play the remaining five games in the order of their preference. Following this, participants will be presented with all 6 games once again in a randomised order. In the final stage, participants will be given free operant choice over which games they wish to play. Six games must be played in total and repeated playing of games will be allowed in this final stage. Participants will be asked to rate each game using the 12-item scale from the DRM (Kahneman et al., 2004).

4.2.1 Method

4.2.1.1 Participants

Thirty participants (16 male, 14 female) all aged 18-25 were recruited from a sample of convenience for the study.

4.2.1.2 Materials

Stimuli consisted of simple coloured geometric shapes, such as green squares, red circles and blue triangles. Stimuli were presented on a computer screen (resolution 1024 x 768) via Microsoft Visual Basic 6.0 software (c.f., Dixon & MacLin, 2003; Cabello, Barnes-Holmes, O'Hora, & Stewart, 2002) and a PC, which also recorded the nature and timing of stimulus presentation and participant responses. Standard computer headphones were used to present auditory stimuli and feedback to participants.

4.2.1.3 Design

Experiment 4.2 employed a 2x3 factorial repeated measures design. The independent variables were the number of different stimuli presented in a level (2-stimuli or 6-stimuli), and the speed at which stimuli were presented in a level (slow, medium, or fast). Thus, there were six game types in total (i.e., 2-stimuli slow, 2-stimuli medium, 2-stimuli fast, 6-stimuli slow, 6-stimuli medium, 6-stimuli fast). There were four stages of the experiment. All participants were initially exposed to all six experimental conditions in a quasi-random order, counterbalanced across participants. In the second stage, participants chose the order in which they played those same games. The third stage consisted of a further presentation of the same six games in a quasi-random order, counterbalanced

across participants. In the fourth stage, participants were given free operant choice over which games they wished to play. Participants were required to play six games in total in the fourth stage, consisting of any combination of the available levels. There were four dependent measures employed; participants' score on each level of the game, their subjectively rated level of both positive and Negative Affect (see Kahneman et al., 2004), and the games chosen in the second and fourth stages of the experiment.

4.2.1.4 Procedure

Participants were first seated in front of the computer monitor and given a brief overview as to the nature of the study. Instructions were given that only the mouse was to be used for interacting with the computer programme. The headphones were then placed over the participants' ears and the volume on the computer was tested to ensure that participants could hear the presented stimuli comfortably. Participants were then directed to read the on-screen instructions and begin the experiment.

A number of instructions were then presented on-screen, one at a time, in a large font. Once each screen had been read, participants could proceed to the next one by pressing the continue button. The instructions presented were as follows, where the text within quotation marks represents the text presented on each individual screen:

"Thank you for agreeing to participate in this research. First, MAKE SURE YOUR PHONE IS SWITCHED OFF !!!!!!! Please click continue to proceed"

"In the next few minutes a number of different games will be presented"

"In these games you will have to gain points by SAVING or DESTROYING the objects which approach you"

"You can SAVE the objects by clicking on them with your mouse pointer"

"You can DESTROY the objects by clicking on the destroy button below"

At this point a graphic representation of the control bar used in the game was displayed, with the DESTROY button pointed out. Participants could proceed to the next instructions by pressing the continue button:

"Please note your score is located in the bottom right hand corner of the screen"

At this point a graphic representation of the control bar used in the game was displayed, with the Score Display pointed out. Participants could proceed to the next instructions by pressing the continue button:

"Before and after each game you will be informed which game you are playing"

"Each game will be described with a speed (i.e., Slow, Medium, Fast)....."

"....And a difficulty level (i.e., Easy, Difficult)"

"For example, a game may be described as FAST and DIFFICULT, while another game may be described as SLOW and EASY"

"Take note of which levels you enjoyed the most"

"Please Click below when you are ready to begin"

Participants were then prompted to click the on-screen, 'continue,' button in order to begin the first game.

Game

The six games presented were almost identical to those presented in Experiment 4.1. For a description of the operation of a typical game and of how these games differ from each other see the procedure section from Experiment 4.1. However, a number of small modifications to the procedure of Experiment 4.1 were made for the purposes of experiment 4.2. Firstly, the stimuli in the slow games in Experiment 4.2 increased in size every 300ms, compared to every 225ms in Experiment 4.1. Thus, each trial in a slow game in Experiment 4.2 lasted a maximum of 2400ms. The medium and fast games were not modified for Experiment 4.2.

Secondly, in order that participants could identify each game when given a choice of games to play, the name of each game was presented for 5000ms at the beginning and end of each game. For example, in the 2-stimuli-slow game, "This game is EASY and SLOW," was presented on-screen for 5000ms.

Thirdly, during the presentation of the questionnaire in Experiment 4.1 each question was presented for ten seconds before the participant could move on to the next question. This was designed to provide a rest period between levels, so that EDA levels could return to baseline. As there was no EDA recording done in Experiment 4.2, these ten second delays were removed and participants could move through the questionnaire as quickly or slowly as they wished.

While the games themselves were very similar across Experiments 4.1 and 4.2, the experimental procedure in which these games were presented was quite different across these experiments, and will be described in detail below. Experiment 4.2 consisted of four distinct stages, which all participants were required to complete.

Stage 1

Stage 1 consisted of the quasi-randomised presentation of all six games. Participants played each of the six games in an order that was determined by the computer programme and was counter-balanced across participants. The 12-item questionnaire was presented after each game.

Stage 2

At the beginning of Stage 2, participants were presented with a screen displaying a ‘continue’ button and the following instructions:

“In the next phase you will be able to choose which game you would like to play.”

Once participants clicked on the ‘continue’ button a game-choice screen was presented with instructions at the top of the page and six buttons in the lower part of the page (as illustrated in figure 4.14). The instructions presented were as follows:

"Please choose which game you would like to play."

Each of the six buttons was labeled with the name of a game. Participants were free to choose any of the games. Once any button was clicked, indicating game choice, that particular game was presented and played by participants. The 12-item questionnaire was then presented, as in Stage 1. Once the questionnaire had been completed, the participant was presented with the game-choice screen. However, the game that had been previously played was now unavailable. Participants were required to choose which of the remaining five games they wished to play. This game was then followed by the presentation of the questionnaire and a further game-choice screen, in which both of the previously played games were unavailable for play. This pattern continued until all six games had been played. This procedure was necessary to ensure that all games were played an equal number of times by all participants in the first three stages of the experiment. Once all six games had been played, participants advanced on to Stage 3.

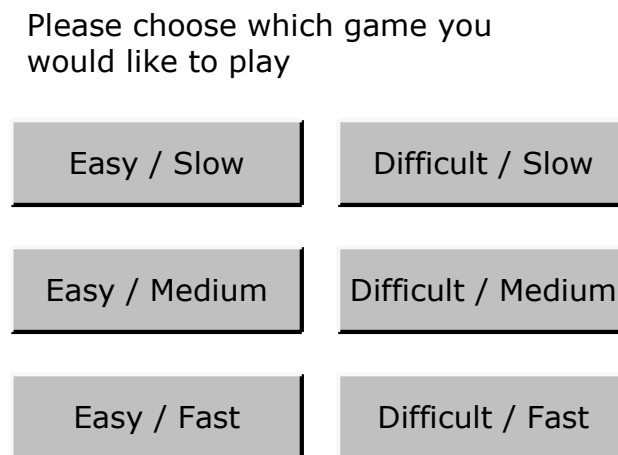


Figure 4.14: Screenshot of the game choice screen presented to participants during Stage 2 and 4 of Experiment 4.2

Stage 3

At the beginning of Stage 3, participants were presented with a screen displaying a ‘continue’ button and the following instructions:

“We will now ask you to play all six games again in a random order. Please try and gain as high a score as possible.”

Identically to Stage 1, Stage 3 consisted of the quasi-randomised presentation of all six games. Participants played each of the six games in an order that was determined by the computer programme and was counter-balanced across participants. The 12-item questionnaire was presented after each game.

Stage 4

At the beginning of Stage 4, participants were presented with a screen displaying a ‘continue’ button and the following instructions:

“In the next phase you will be able to choose which game you would like to play. You will be given this choice a number of times”

As in Stage 2, once participants clicked on the ‘continue’ button a game-choice screen was presented with instructions at the top of the page and six buttons in the lower part of the page (as illustrated in figure 4.14). The instructions presented were as follows:

"Please choose which game you would like to play."

Each of the six buttons was labeled with the name of a game. Participants were free to choose any of the games. Once any button was clicked, indicating game choice, that particular game was presented and played by participants. The 12-item questionnaire was then presented, as in Stage 1. Once the questionnaire had been completed, the participant was presented with the game-choice screen. Unlike in Stage 2, all games were available to be chosen again. Participants were required to play six games in total. Thus, Stage 4 consisted of six consecutive free-operant choices between all of the games previously presented.

4.2.2 Results and Discussion

Thirty participants (16 male, 14 female) played the full sequence of games. In Stage 2, participants were given the opportunities to choose to play any of the six available games. For the remainder of this phase choices between games did not include previously played games. Because the first choice made is being taken to represent participants' preferred game, choices following the first choice in Stage 2 are not analysed here. However, the sequence of choices made by participants can be seen in Appendix 3. In Stage 4, repeated play of games was allowed and thus participants' choices of game represented their preferred game on every play. Thus, the entire sequences of choices made in Stage 4 were subjected to analysis and can be seen below.

Figure 4.15 displays the number of participants who chose each of the six games, during phase 2 and 4. The first panel to the left of the graph represents the number of participants who chose each game at the first opportunity in Stage 2. The remaining six panels to the right of the graph represent sequentially each of the six choices made in Stage 4 of the experiment.

On the first choice (Stage 2), more participants chose to play the 2-stimuli-fast game than any other game. The 6-stimuli-slow and 6-stimuli-fast games were chosen less times than any other game. On the second choice (Stage 4), more participants chose the 2-stimuli-slow game and 6-stimuli-medium game than any other game. The 6-stimuli-fast game was chosen by fewer participants than any other game. On the third choice, the 6-stimuli-medium and 6-stimuli-fast games were the most popular, while the 2-stimuli-fast game was the least popular. On the fourth choice, the 6-stimuli-medium game was the most popular, while the 2-stimuli-slow game was the least popular. On the fifth choice, the 6-stimuli-fast game was chosen by more participants than any other game, while the 2-stimuli-slow and 6-stimuli-slow games were the least popular. On the sixth choice, the 6-stimuli-medium and six stimuli fast games were the most popular, while the 2-stimuli-medium game was the least popular. On the seventh and final choice, the

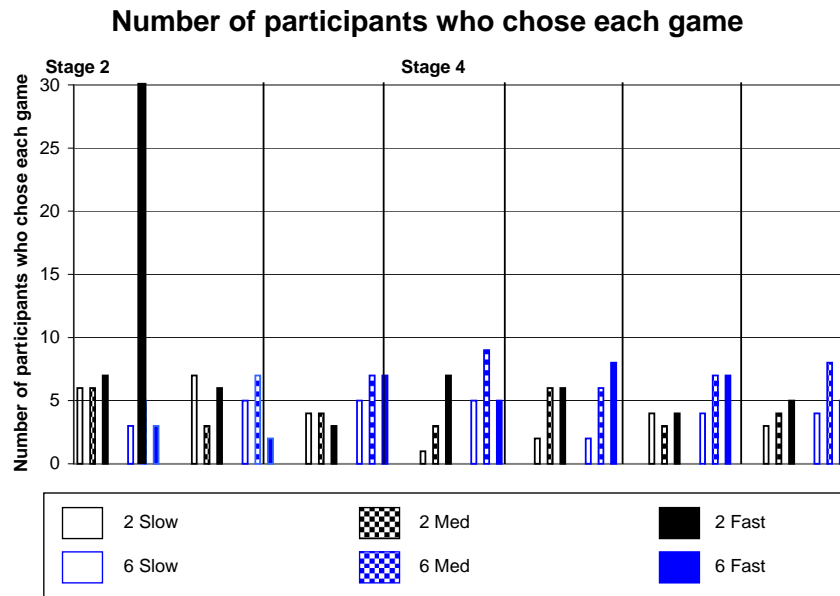


Figure 4.15: Graphic representation of the number of participants who chose each game during Stages 2 and 4

6-stimuli-medium game was chosen by more participants than any other game, while the 2-stimuli-slow game was the least popular. The 6-stimuli-medium game was the most popular game across all opportunities to choose, while the 2-stimuli-slow game was the least popular.

Interestingly, there appears to be a shift in preference from the simpler games to the more complex games across presentations. For example, the 2-stimuli-slow game was the second most popular game when participants were first presented with an opportunity to choose, and was the most popular game on their second opportunity. However, by the fourth opportunity to choose which game they wished to play, participants had begun to rate the 2-stimuli-slow game as the least popular game. Moreover, the 2-stimuli-slow game remained one of the least popular games across the remaining choices. Conversely, the 6-stimuli-fast game was the least popular game at the start of Stage 2 and the start of Stage 4. By the second opportunity to choose in Stage 4, the 6-stimuli-fast game was the most popular game and remained one of the most popular games across the remaining choices. Thus, the free-operant measure used here allowed the experimenter to observe that upon repeated presentations, preferences shift from simpler games to more complex games.

Figures 4.16-4.18 present data for mean Correct Responses, Positive Affect and Negative Affect across all participants for the six games in each of the first three stages in the current experiment. Data for stage four is not presented, as different numbers of participants played each game in this stage, rendering an unbiased comparison between games impossible. Correct Responses refers to the

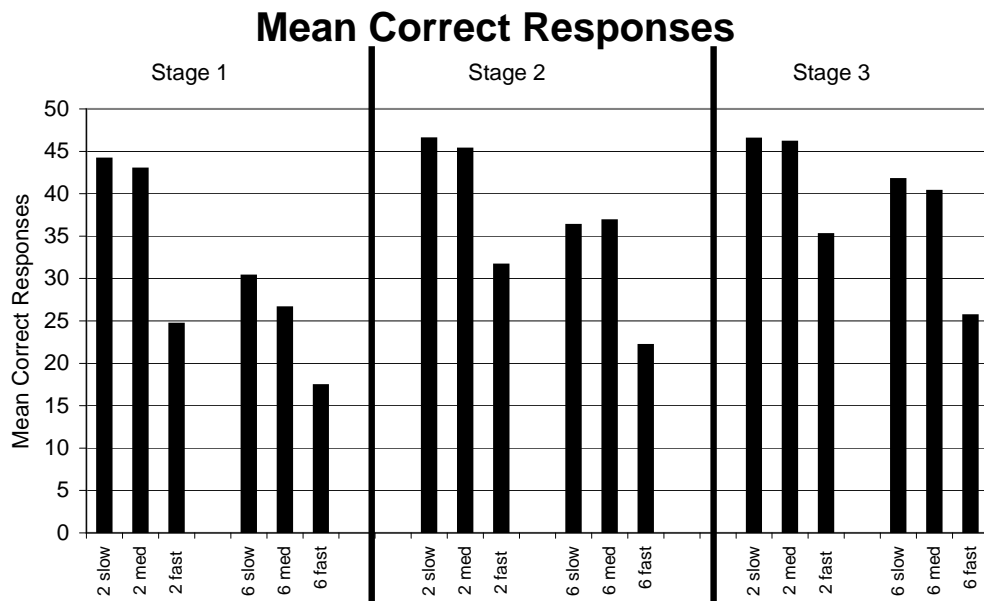


Figure 4.16: Mean Correct Responses for each game across stages 1-3 (n=30)

total number of Correct Responses participants made to the stimuli presented in a particular level. The 12 item DRM questionnaire, which was presented to participants after each game, allowed the experimenter to calculate Negative Affect and Positive Affect scores for each game level. Participants rated whether they agreed or disagreed with a number of statements on a six point sliding scale. Six of the questions referred to negative experiences, so the answers to these six questions were summed to form the statistic Negative Affect (i.e., maximum total score = 36). Three of the remaining questions referred to positive experiences and the answers to these three questions were summed to form the statistic Positive Affect (i.e., maximum total score = 18).

Figure 4.16 displays mean Correct Responses for each level across the first three stages of the experiment. The reader is reminded that Stages 1 and 3 involved the presentation of all six games in a randomised order, while in Stage 2 participants chose the order in which they played those six games. In the first stage, mean Correct Responses were highest in the 2-stimuli-slow game and lowest in the 6-stimuli-fast game. In the second stage, mean Correct Responses were highest in the 2-stimuli-slow game and lowest in the 6-stimuli-fast game. In the third stage, mean Correct Responses were again highest in the 2-stimuli-slow game and lowest in the 6-stimuli-fast game.

While the pattern of which game produced the highest and lowest number of Correct Responses is consistent across all three stages, the absolute mean scores obtained by participants did change across repeated presentations. Correct Responses for the 2-stimuli-slow and 2-stimuli-medium games

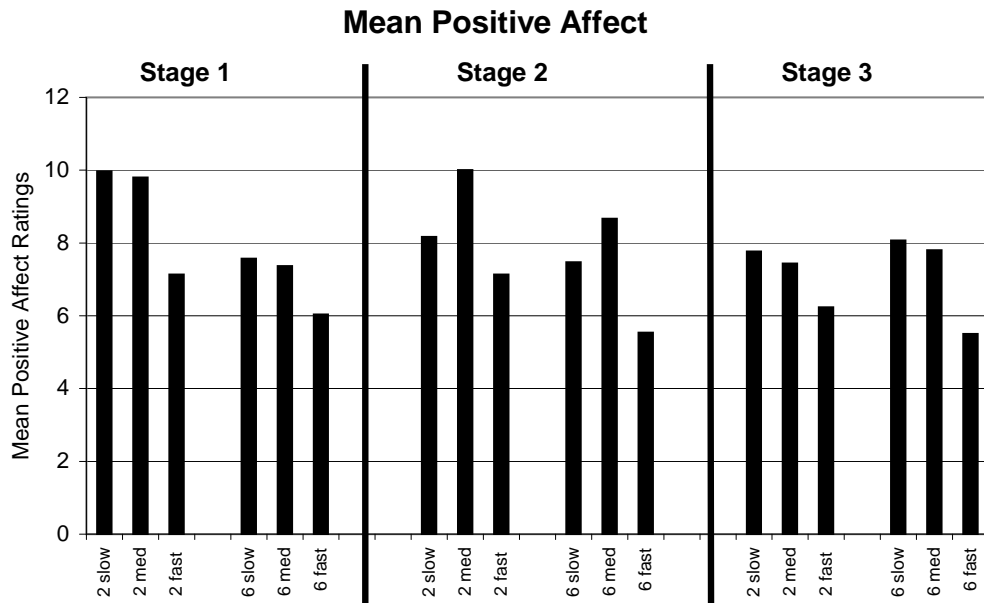


Figure 4.17: Mean Positive Affect ratings for each game across stages 1-3 (n=30)

were quite high on the first presentation and maintained at that high level across the following two stages. In effect, it appears that there was a ceiling effect on these scores (i.e., there were only 48 trials in every game). However, mean Correct Responses on all four other games increased dramatically across the three stages of the experiment. For example, participants, on average, produced 25 Correct Responses on the 6-stimuli-medium game on their first exposure to this game. On their second exposure to this game, participants produced approximately 37 Correct Responses, while they produced approximately 40 on their third exposure. In effect, it appears that participants' accuracy and fluency in responding on the 6 stimuli games approached that observed in the 2 stimuli games, across repeated presentations of all six games. However, it must be noted that while scores on the fast games did approach those registered on the slow games across successive presentations, they were still relatively low on the third presentation.

Figure 4.17 displays mean Positive Affect ratings for each game across the first three stages of the experiment. The reader is reminded that stages one and three involved the presentation of all six games in a quasi-randomised order that was counter-balanced across all participants. In Stage 2, participants were participants made a series of choices as to which game they wished to play next. Each time a choice was made, that game was then unavailable to be chosen later.

In the first stage, Positive Affect ratings were highest in the 2-stimuli-slow game and lowest in the 6-stimuli-fast game. In the second stage, Positive Affect ratings were highest in the

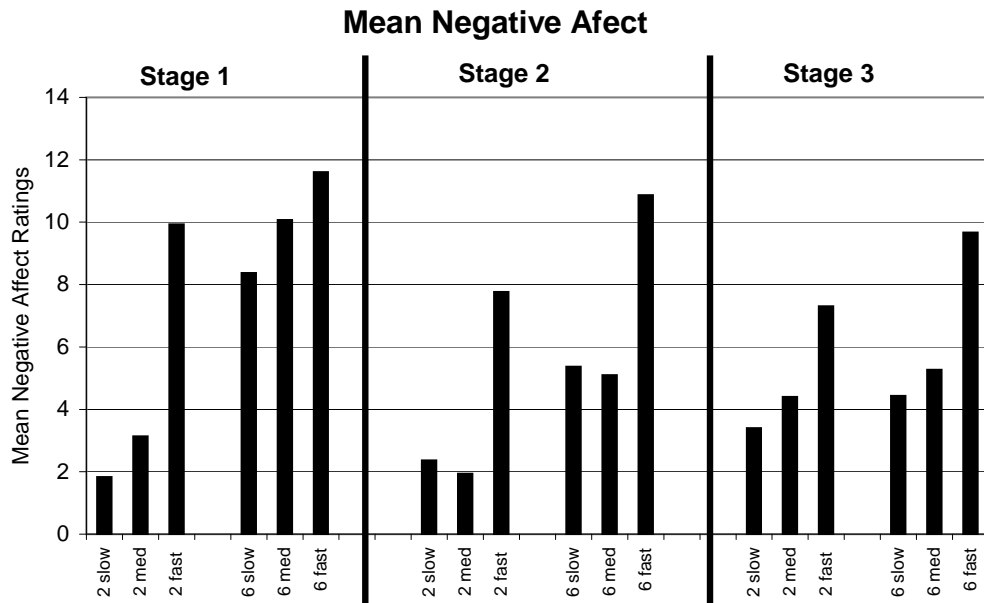


Figure 4.18: Mean Negative Affect ratings for each game across stages 1-3 (n=30)

2-stimuli-medium game and lowest in the 6-stimuli-fast game. In the third stage, Positive Affect ratings were highest in the 6-stimuli-slow game and lowest in the 6-stimuli-fast game. Thus, while no game was consistently rated as the highest in terms of Positive Affect, the 6-stimuli-fast game was consistently rated as the lowest.

Mean Positive Affect ratings appear to be relatively consistent across the first three stages of the experiment, for most of the games presented. However, the 2-stimuli-slow game and 2-stimuli-medium game do show a distinct change across consecutive presentations. The 2-stimuli-slow game is rated relatively high in terms of Positive Affect on the first presentation, but is then rated similar to the rest of the levels on the second and third presentations. The 2-stimuli-medium game is rated relatively high in terms of Positive Affect on the first and second presentations, but is then rated as similar to the rest of the levels on the third presentation. It appears that, while the simpler games were rated as higher in Positive Affect upon the first presentation, participants rated games as being more similar to each other in terms of Positive Affect across repeated presentations.

Figure 4.18 displays mean Negative Affect ratings for each game across the first three stages of the experiment. In the first stage, Negative Affect ratings were highest in the 6-stimuli-fast game and lowest in the 2-stimuli-slow game. In the second stage, Negative Affect ratings were highest in the 6-stimuli-fast game and lowest in the 2-stimuli-medium game. In Stage 2, Negative Affect

ratings were highest in the 6-stimuli-fast game and lowest in the 2-stimuli-slow game. The 6-stimuli-fast game appears to be consistently rated as the most negative, while the two stimuli slow and 2-stimuli-medium games appear to be the least negative.

Mean Negative Affect ratings for the 2-stimuli-slow and 2-stimuli-medium games appear to be relatively consistent across the first and second presentations. However, there was a small rise in ratings for both of these levels, from the second to third presentations. Mean Negative Affect ratings for the 6-stimuli-medium game decrease across the first and second presentations, before increasing marginally from the second to third presentations. Mean Negative Affect ratings for the remaining three games (i.e., 2-stimuli-fast, 6-stimuli-slow and 6-stimuli-fast) decreased linearly across repeated presentations. Thus, it appears that games that were initially rated as high in Negative Affect are rated lower upon repeated presentations, while games that were initially rated as low in Negative Affect are rated marginally higher across repeated presentations.

Ratings of all levels appear to be converging across time. This pattern is consistent with the pattern observed with Positive Affect. Thus, a general trend is apparent in which the simpler games are being rated as less positive and more negative across successive presentations, while the more complex games appear to be rated as more positive and less negative across successive presentations.

As mentioned in the introduction to the current chapter, doubts remain over the utility of using subjective verbal reports as a measure of enjoyment in a behaviour analytic study. In order to assess the validity of the 12 item scale used in the current study as a measure of enjoyment in games, participants' ratings of games in terms of positive and Negative Affect were correlated with their subsequent choice of which game they wished to play again. Specifically, ratings from Stage 1 were considered in relation to participants' choice of which game they wished to play at the beginning of Stage 2. Additionally, ratings from Stage 3 were considered in relation to participants' choice of which game they wished to play at the beginning of Stage 4.

Examining ratings of Positive Affect first, the game that was rated highest or joint highest in terms of Positive Affect in Stage 1 was chosen by 10/30 participants as the game they wished to play at the beginning of Stage 2. Subsequently, in only 11/30 cases was the game that was rated highest in Positive Affect in Stage 3 chosen by participants as the game they wished to play at the beginning of Stage 4. It appears that ratings of Positive Affect predicted actual game preference in approximately 33% of cases. Examining ratings of Negative Affect, the game that was rated lowest or joint lowest in terms of Negative Affect in Stage 1 was chosen by 11/30 participants as the game they wished to play at the beginning of Stage 2. Subsequently, in only 6/30 cases was the game that was rated lowest in Negative Affect in Stage 3 chosen by participants as the game they wished to

play at the beginning of Stage 4. It appears that ratings of Negative Affect predicted actual game preference in approximately 26% of cases. Correlations of 33% and 26% respectively, between ratings of positive and Negative Affect and subsequently observed behaviour, are very low and suggest that the 12 item scale employed in the current study was not a useful tool for evaluating participants' enjoyment of games.

The current study aimed to employ a novel behavioural measure for assessing enjoyment in games and also to evaluate the 12-item DRM questionnaire as a tool for measuring enjoyment in games. The behavioural measure involved allowing participants to choose which game they wanted to play. The choice that participants made necessarily defined that participants' favourite (i.e., most reinforcing) game. Thus, it was possible to compare ratings of Positive and Negative Affect with these choices, in order to evaluate the validity of the questionnaire. It was found that ratings of Positive Affect predicted actual game preference (i.e., game choice) in approximately 33% of cases, while ratings of Negative Affect predicted game choice in approximately 26% of cases. That verbal ratings have fared so poorly as predictors of overt behaviour in the current experiment is not surprising, given the widely reported poor relationship between verbal reports and overt behaviour (see Cabello & O'Hora, 2002; Critchfield & Epting, 1998; Dymond & Barnes, 1997; Perone, 1988, for discussion). Thus, it appears that the free operant measure would appear to have the best face validity in terms of measuring enjoyment/preference in the current experiment.

Interestingly, results from Stage 1 of Experiment 4.2 are very similar to those recorded in Experiment 4.1, and those reported in a number of experiments conducted in previous chapters. These experiments all found that participants scored significantly better on, and displayed preference for, simple games over more complex games. Thus, the current study qualifies the findings of these previous studies. Importantly, however, Experiment 4.2 also expands on these findings, in that the analysis demonstrated how preference this preference towards simpler games changes across repeated presentations. Specifically, preference for complex games, as measured by both self-report and free-operant measures, increased across presentations of those games. This finding sits well with the literature on 'challenge' in gaming, which suggests that players seek out challenges appropriate to their own skill level (i.e., Gingold, 2005; Malone, 1982; Morlock et al., 1985; Vorderer et al., 2003). While results of previous studies merely suggested that players prefer simpler games, the current study refined this definition to take account of the effect of experience. It is conceivable that such a shift in preference towards the more complex games may have been observed in the previous studies, if a similar procedure had been employed.

Experiments 4.1 and 4.2 have each established a novel potential measure of the reinforcing

properties (i.e., enjoyment) of computer games. Thus, these two measures would appear to warrant further investigation. Moreover, additional research is needed to examine these separate measures in relation to each other. The following experiment addresses these issues.

4.3 Experiment 4.3

Experiments 4.1 and 4.2 have each established a novel potential measure of the reinforcing properties (i.e., enjoyment) of computer games. Specifically, Experiment 4.1 demonstrated that it may be possible to discriminate whether a player is enjoying a particular game from an ongoing analysis of that game players' Skin Conductance Levels. This finding was made by comparing the slopes of SCL data during games that participants reported as enjoyable, to those they reported as not enjoyable. Experiment 4.2 employed a novel behavioural measure for assessing the reinforcing properties of games. Importantly, the behavioural measure employed in Experiment 4.2 was also used to evaluate the face validity of the self-report questionnaire used in the majority of studies in the current research programme. This analysis revealed that the self-report questionnaire was not a useful tool for predicting game preference (i.e., choice). Thus, as the SCL measure developed in Experiment 4.1 was evaluated through correlating SCL trends with self-reports, further work must be conducted in order to assess the utility of the SCL measure in predicting behaviour.

Experiment 4.3 will involve combining the two novel measures of the reinforcing properties (i.e., enjoyment) of computer games developed in the previous two experiments. Specifically, Skin Conductance Levels will be recorded as participants play the six games that were presented in Experiments 4.1 and 4.2. Once all six games have been played, participants will be given a choice of which game they wish to play again. When they finish playing this chosen game, the experiment is finished. SCL data from games that participants choose to play will be analysed in order to identify any characteristic trends, and compared to SCL data from games that are not chosen. After each of the seven games, participants will also be asked to rate that level using the 12-item scale from the DRM (Kahneman et al., 2004). This measure is included to maintain as much procedural consistency across the experiments in Chapter 4 as possible, to allow comparison across those three experiments, and to support and expand upon the findings of Experiment 4.1.

It must be noted that this procedure reverts to the presentation of each game once only. It was pointed out in Experiment 4.2 that such a preparation necessarily leads participants to show preference for the easier games. Thus, the procedure of the current experiment is knowingly biasing participants towards favouring the simpler games. However, Experiment 4.3 is not concerned with

evaluating which game participants enjoyed most. This experiment is designed to examine SCL correlates of the game which participants most enjoyed on the first presentation (i.e., the one chosen subsequently to be played again).

4.3.1 Method

4.3.1.1 Participants

Eleven participants, (7 male, 4 female) all undergraduate or postgraduate university students, aged 18-23, were recruited from a sample of convenience.

4.3.1.2 Materials

Materials used were identical to those used in the preparation of Experiment 4.1, with the exception that there were some minor alterations to the Visual Basic 6.0 programme used to present stimuli and record responses, as discussed in the procedure section.

4.3.1.3 Design

Experiment 4.3 employed a 2X3 factorial repeated measures design. The independent variables were the number of different stimuli presented in a level (2-stimuli or 6-stimuli), and the speed at which stimuli were presented in a level (slow, medium, or fast). Thus, there were six game types in total (i.e., 2-stimuli slow, 2-stimuli medium, 2-stimuli fast, 6-stimuli slow, 6-stimuli medium, 6-stimuli fast) In stage 1, all participants were initially exposed to all six games in a counterbalanced order. In addition to the four dependent measures employed in Experiment 4.1, (SCL, Correct Responses, Positive Affect and Negative Affect) an operant measure was employed in Experiment 4.2 that indicated which game each participant chose to play in Stage 2 of the experiment.

4.3.1.4 Procedure

Experiment 4.3 consisted of two stages. Stage 1 essentially consisted of the procedure employed in Experiment 4.1. Stage 2 presented participants with a free-operant choice, similar to the procedure employed in Experiment 4.2.

Stage 1

Stage 1 of Experiment 4.3 was essentially identical to the procedure of Experiment 4.1, apart from two small modifications, which were employed in Experiment 4.2. Firstly, the stimuli in the

slow games in Experiment 4.3 increased in size every 300ms, compared to every 225ms in Experiment 4.1. Thus, each trial in a slow game in Experiment 4.3 lasted a maximum of 2400ms. The medium and fast games were not modified for Experiment 4.3.

Secondly, in order that participants could identify each game when given a choice of games to play, the name of each game was presented for 5000ms at the beginning and end of each game. For example, in the 2-stimuli-slow game, "This game is EASY and SLOW," was presented on-screen for 5000ms. Once participants had completed Stage 1, they were presented with Stage 2.

Stage 2

At the beginning of Stage 2, participants were presented with a screen displaying a 'continue' button and the following instructions:

"In the next phase you will be able to choose which game you would like to play."

Once participants clicked on the 'continue' button a game-choice screen was presented with instructions at the top of the page and six buttons in the lower part of the page (as illustrated in Figure 4.14, Experiment 4.2). The instructions presented were as follows:

"Please choose which game you would like to play."

Each of the six buttons was labeled with the name of a game. Participants were free to choose any of the games. Once any button was clicked, indicating game choice, that particular game was presented and played by participants. The 12-item questionnaire was then presented, as in Stage 1. Once the questionnaire had been completed, the experiment was finished and participants were de-briefed.

4.3.2 Results and Discussion

Eleven participants (7 male, 4 female) completed the entire experimental procedure. Upon initial data analysis it was observed that EDA data was not successfully recorded from two participants (1 male, 1 female). Thus, the following data analysis was conducted on the remaining nine participants. Figure 4.19 presents the number of participants who chose to play each of the six available games in Stage 2 of Experiment 4.3.

Figure 4.19 displays the number of participants who chose to play each game in Stage 2 of Experiment 4.3. The 6-stimuli-medium speed game was chosen by more participants than any other game. The 6-stimuli-fast game was the least popular. The 2-stimuli-medium and 2-stimuli-fast games were more popular than the 2-stimuli-slow and 6-stimuli-slow games. Participants did not

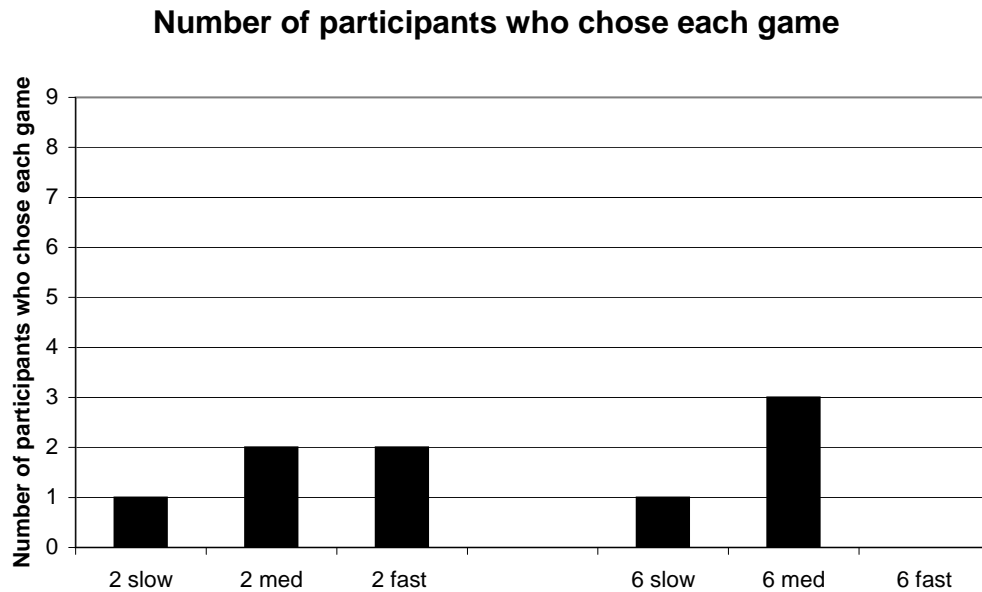


Figure 4.19: The number of participants who chose to play each game in Stage 2 of Experiment 4.3

show preference for either the easiest (i.e., 2-stimuli-slow) or most complex (i.e., 6-stimuli-fast) games. Indeed, preference was spread quite evenly across five of the six games. As a small number of participants chose a wide range of games, it is unlikely that stable trends will emerge in either self-reports of enjoyment, or SCL data, for any one game

Table 4.6 and figures 4.20-4.22 present mean data for Correct Responses, Positive Affect and Negative Affect across all participants for the six games employed in Stage 1 of the current experiment. Correct Responses refers to the total number of Correct Responses participants made to the stimuli presented in a particular level. The 12 item DRM questionnaire, which was presented to participants after each game, allowed the experimenter to calculate Negative Affect and Positive Affect scores for each game. Participants rated whether they agreed or disagreed with a number of statements on a six point sliding scale. Six of the questions referred to negative experiences, so the answers to these six questions were summed to form the statistic Negative Affect (i.e., maximum total score = 36). Three of the remaining questions referred to positive experiences and the answers to these three questions were summed to form the statistic Positive Affect (i.e., maximum total score = 18).

Table 4.6 presents mean values for Correct Responses, Positive Affect and Negative Affect, for all games employed in Stage 1 of Experiment 3. For ease of analysis these data points are represented below in figures figures 4.20-4.22.

Figure 4.20 illustrates that the 2-stimuli-slow game produced the highest mean number of

Table 4.6: Mean Correct Responses, Positive Affect and Negative Affect across all six games played in Stage 1 (n=9)

	Slow	Medium	Fast
Mean Correct Responses			
2-Stimuli	45.889	42.778	25.444
6-Stimuli	31.444	28.000	14.889
Mean Positive Affect			
2-Stimuli	10.333	11.000	5.889
6-Stimuli	8.111	8.444	4.444
Mean Negative Affect			
2-Stimuli	4.667	4.889	15.444
6-Stimuli	7.111	13.222	15.111

Mean Correct Responses

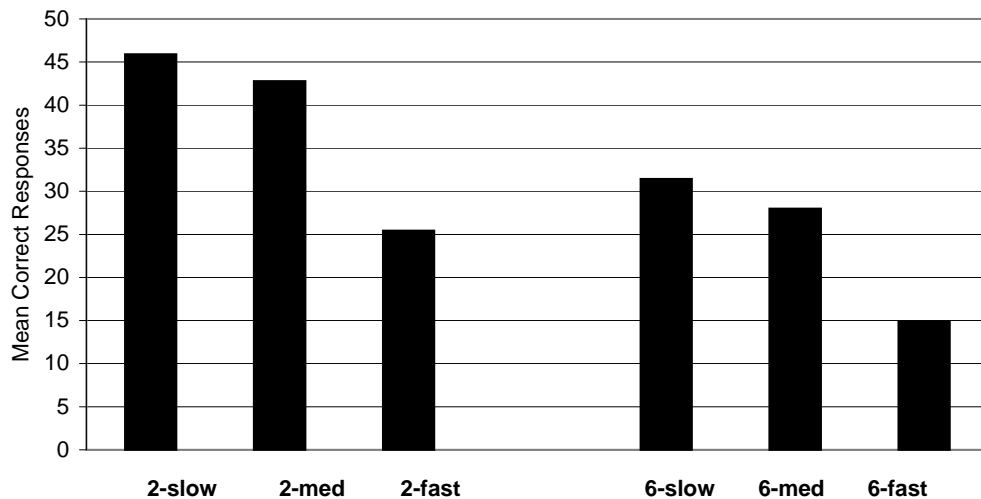


Figure 4.20: Mean Correct Responses across the six games played in Stage 1 (n=9)

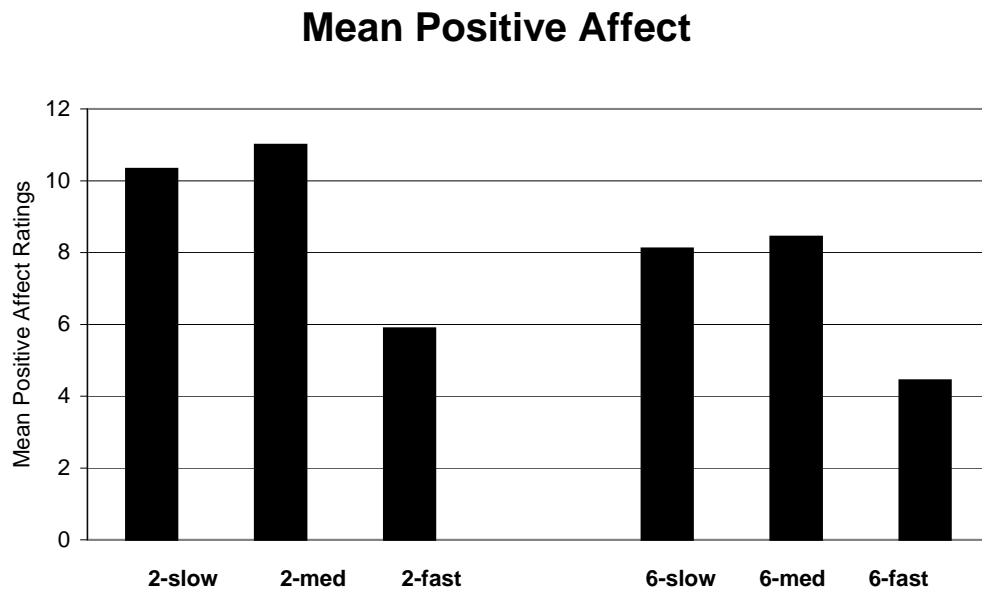


Figure 4.21: Mean Positive Affect ratings across the six games played in Stage 1 (n=9)

Correct Responses, while the 6-stimuli-fast game produced the lowest. Mean Correct Responses are generally higher in the games involving the presentation of 2-stimuli than the 6-stimuli games, which suggests that participants were less successful when playing a game involving a larger amount of stimuli. However, this was not true in all cases. Indeed, the 2-stimuli-fast game produced a lower mean number of Correct Responses than both the 6-stimuli-slow and 6-stimuli-medium games. There is a linear pattern of lower mean Correct Responses in the faster levels. In both the 2-stimuli and 6-stimuli games there are lower mean Correct Responses in the medium speed game than the slow game, and lower mean Correct Responses again in the fast game over the medium speed game.

Figure 4.21 illustrates that the 2-stimuli-medium game produced the highest mean Positive Affect rating, while the 6-stimuli-fast game produced the lowest mean Positive Affect rating. Mean Positive Affect ratings were generally higher in the games involving the presentation of two stimuli, suggesting that participants preferred the 2-stimuli games. However, this was not true in all cases. The 2-fast game produced lower ratings of Positive Affect than the 6-stimuli slow and 6-stimuli-medium games. Interestingly, there is an ‘n’ shape to Positive Affect ratings across speeds. Specifically, across speeds in both the 2-stimuli and 6-stimuli games, ratings of Positive Affect are highest for the medium speed games and lowest for the fast games. Thus, unlike in previous experiments, there is no linear relationship of lower Positive Affect reported as speed increases. Indeed, the medium speed games appear to be preferred by participants in Experiment 4.3, as measured by both the self-report questionnaire and the overt behavioural measure.

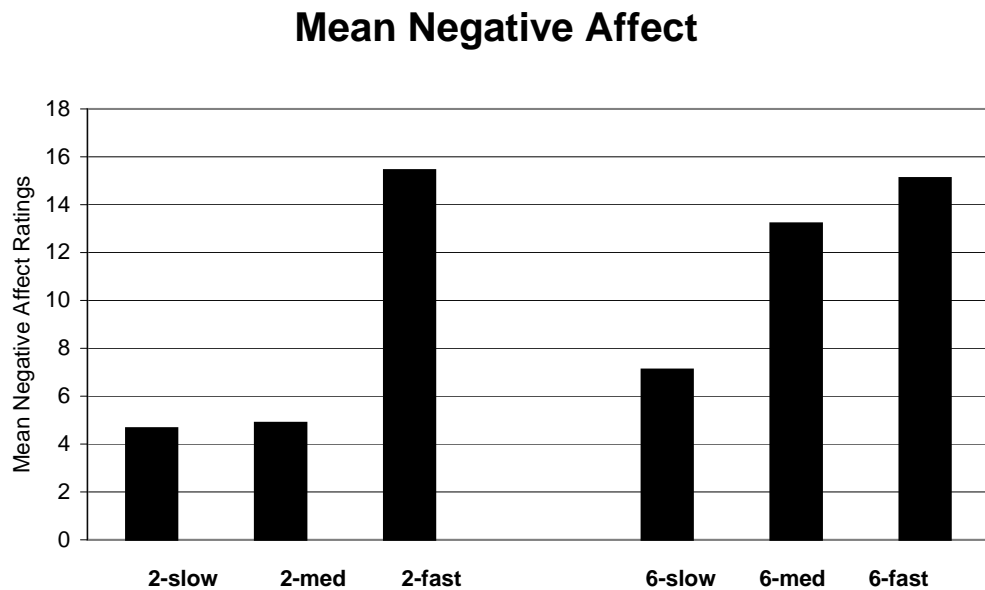


Figure 4.22: Mean Negative Affect ratings across the six games played in Stage 1 (n=9)

Figure 4.22 illustrates that the 2-stimuli-fast game produced the highest mean Negative Affect rating, while the 2-stimuli-slow game produced the lowest mean Negative Affect rating. Ratings of Negative Affect are higher for all of the 6 stimuli games than for the 2-stimuli-slow and 2-stimuli-medium games. However, the 6-stimuli games were not rated higher for Negative Affect than the 2 stimuli games in all cases. The 2-stimuli fast game produced higher ratings of Negative Affect than all the 6 stimuli games. Thus, there does not appear to be a preference for either 2-stimuli or 6-stimuli games.

There is a linear relationship of higher mean Negative Affect ratings to the faster games in the 6-stimuli games only. In the 6-stimuli games there are higher mean Negative Affect ratings in the medium speed game than the slow game, and higher mean Negative Affect ratings again in the fast game over the medium speed game. This pattern is not reflected across the 2-stimuli games, where the slow and medium games are rated similarly in terms of Negative Affect while the 2-stimuli-fast game was rated dramatically higher than both of these games.

Data across the Positive Affect and Negative Affect graphs appear somewhat chaotic, in that they do not mirror each other. While an ‘n’ shaped graph was found across speeds on Positive Affect, where the medium games were given highest PA ratings, this was not reflected as a ‘u’ shaped graph in Negative Affect ratings. Additionally, the 2-stimuli-fast game was rated as the highest in terms of Negative Affect, but was not rated as the lowest in terms of Positive Affect. This finding may suggest that the self-report questionnaire did not function as a reliable measure of

enjoyment in the context of the current experiment, which featured low numbers of participants.

There does appear to be a general correlation between subjective ratings and Correct Responses. As found in previous studies, the games in which participants produced the highest number of Correct Responses (i.e., 2-stimuli-slow and 2-stimuli-medium) were the games rated highest in terms of Positive Affect and lowest in terms of Negative Affect. Conversely, the games in which participants produced the lowest number of Correct Responses (2-stimuli-fast and 6-stimuli-fast) were also the games rated as lowest in Positive Affect and highest in Negative Affect. It appears that participants rated games in which they produced a large number of incorrect responses as being less enjoyable than those on which they produced mainly Correct Responses.

In order to assess the validity of the 12 item scale used in the current study as a measure of enjoyment in games, participants' ratings of games in terms of positive and Negative Affect were considered in relation to their subsequent choice of game. Specifically, ratings from Stage 1 were considered in relation to participants' choice of which game they wished to play in Stage 2.

Examining ratings of Positive Affect first, the game that was rated highest or joint highest in terms of Positive Affect in Stage 1 was chosen by 2/9 participants as the game they wished to play in Stage 2. It appears that ratings of Positive Affect predicted actual game preference in approximately 22% of cases. Examining ratings of Negative Affect, the game that was rated lowest or joint lowest in terms of Negative Affect in Stage 1 was chosen by none of the participants as the game they wished to play in Stage 2. It appears that ratings of Negative Affect predicted actual game preference in 0% of cases. Correlations of 22% and 0% respectively, between ratings of positive and Negative Affect and subsequently observed behaviour, are very low and suggest that the 12 item scale employed in the current study was not a useful tool for evaluating participants' enjoyment of games.

Skin Conductance was recorded continuously at a rate of 125 samples per second from the beginning of the experiment until it had been completed. Stimulus markers were sent to the polygraph to signal the onset and conclusion of each individual level of the game. Using these stimulus markers, the segments of raw EDA data that corresponded in time to the game levels were then extracted and exported to a spreadsheet for analysis, using the BrainVision © Analyser software. As the area of the electrodes used in the current study were 0.19634954cm^2 , all EDA data first had to be transformed so that results could be presented in units of microsiemens (μS)/ cm^2 . This was carried out by multiplying all raw EDA data by 5.092958 (i.e., $1/0.19634954$).

Due to the finding that game preference, as measured by game choice in Stage 2 of the current experiment, was spread quite evenly across five of the six games, it appears unlikely that any stable trends would emerge from analysing SCL data across participants for each level, as was conducted in

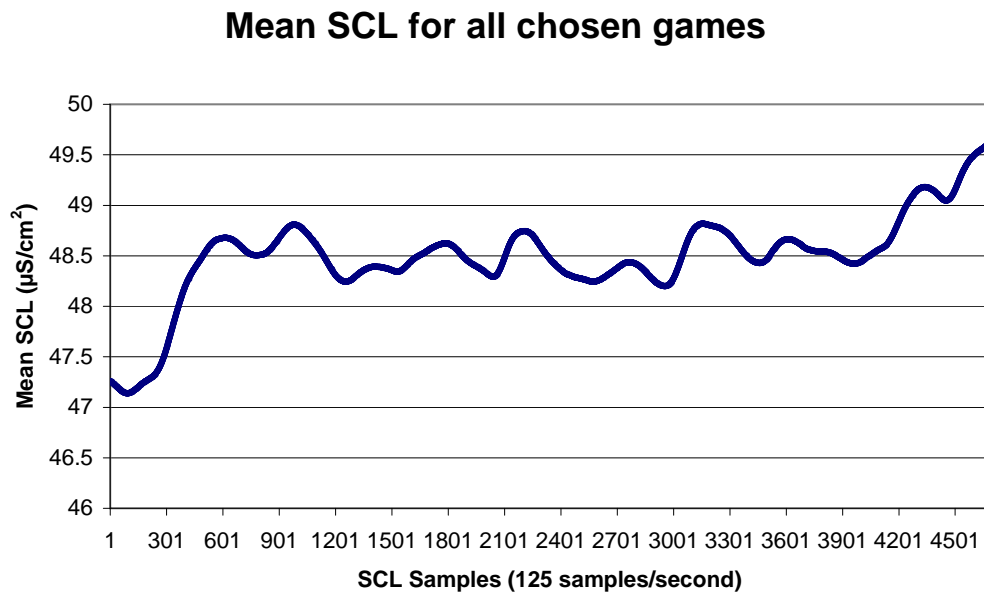


Figure 4.23: Mean SCL recordings for all games in Stage 1 that were later chosen to be played by each participant in Stage 2 (n=9)

Experiment 1. Instead, a choice-based analysis of the SCL data from Stage 1 was carried out. Specifically, SCL data from the game in Stage 1, which participants later chose to play in Stage 2, was identified. This data was then compared with the SCL data from all five games that participants did not choose to play again, in order to evaluate whether any distinctive patterns of change in SCL were observed. This analysis was carried out and the results are presented in figures 4.23 and 4.24. Figure 4.23 presents mean SCL data recorded across all games played in Stage 1 of the experiment, that were later chosen to be played in Stage 2. Figure 4.24 presents mean SCL data recorded across all games played in Stage 1 of the experiment, that were not chosen to be played in Stage 2. Neither of these graphs represent data from any one particular game, as different participants chose to play different games (see figure 4.19 above).

Each participant chose to play one of the six games in Stage 2, thus did not choose to play the other five. As such, while figure 4.23 represents the mean SCL data for the nine chosen games, figure 4.24 represents the mean SCL data from the remaining 45 games played by all participants in Stage 1 of the experiment. In addition, as the duration of SCL recordings differed across games, and across participants, the sessions displayed were cut to the length of the shortest recording. Thus, figures 4.23 and 4.24 present data for the first 4678 samples (or 37.424 seconds) of each session only.

Figure 4.23 presents Mean SCL across all participants in Stage 1, for the game that each participant later chose to play in Stage 2 of the experiment. Mean SCL increased sharply over the

Mean SCL for all games not chosen

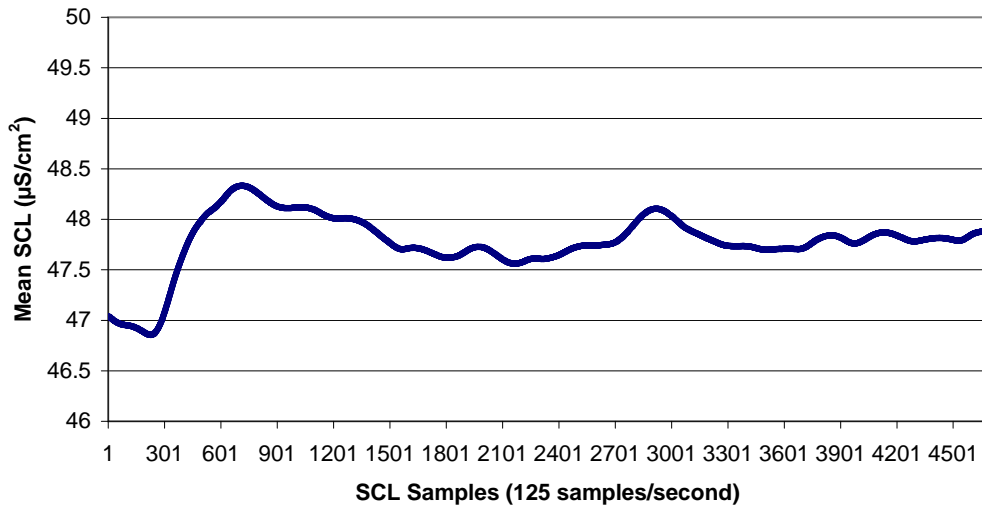


Figure 4.24: Mean SCL from every game in Stage 1 that was not chosen to be played in Stage 2 (n=9)

first five seconds or so, before stabilizing for a period, followed by a further increase towards the end of the session.

Figure 4.24 presents Mean SCL across all participants in Stage 1, for every game that participants did not choose to play in Stage 2 of the experiment. Mean SCL increased sharply over approximately the first five seconds, before decreasing gradually over a period. This was followed by a period where SCL levels remained consistent until the end of the session.

Two clear patterns are apparent across the games which were chosen to be played later, and those that were not. During the games that were later chosen to be played again, Mean SCL increased sharply, and then maintained that level before demonstrating a gradual increase towards the end of the session. Mean SCL across all of the games that were not chosen later also showed a sharp initial increase, but this was followed by a gradual decrease in SCL, before a period where SCL remained consistent. This finding supports that of Experiment 1, where it was proposed that a gradual increase in SCL may be indicative of participants enjoying a game, while any gradual decrease in SCL may be indicative of games that were not enjoyed.

In order to further illustrate the difference in SCL data between games that were chosen and those that were not, best-fit lines were created for the graphs presented in figures 4.23 and 4.24 and the slope and correlation coefficients for these lines were calculated and presented in figure 4.25 and table 4.7. As carried out in Experiment 4.1, the first five seconds of all EDA recordings were

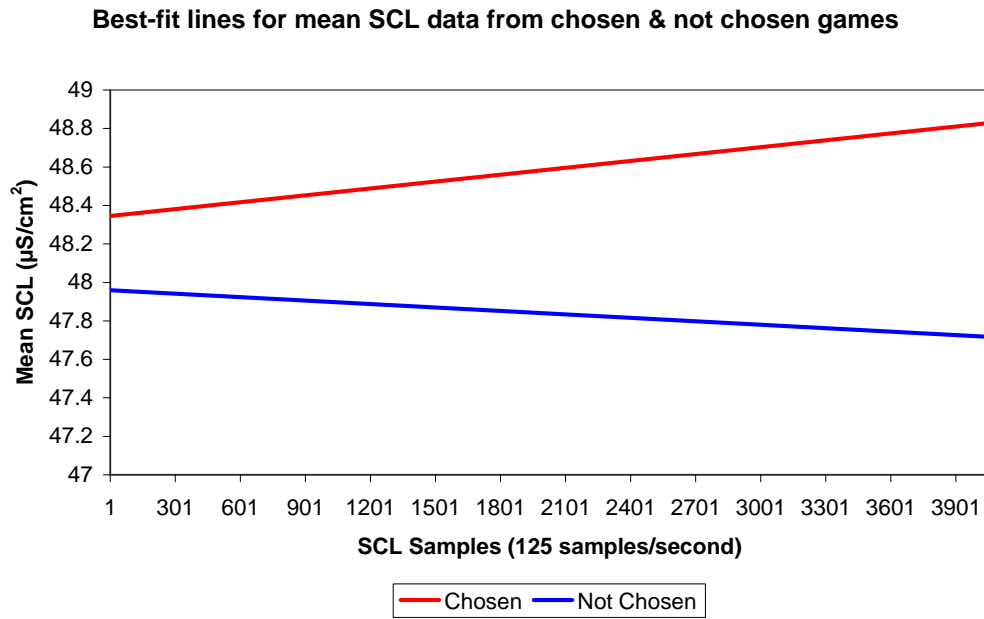


Figure 4.25: Best-fit line for mean SCL recordings for all games in Stage 1 that were later chosen to be played by each participant in Stage 2, and from those that were not chosen (n=9)

Table 4.7: Slope and correlation coefficients (r^2 . values) for games in Stage 1 that were later chosen to be played in Stage 2 and from those that were not chosen to be played in Stage 2 (n=9)

	Slope ($\mu\text{S}/\text{cm}^2$)/S	r^2 . value
Chosen Games	0.0125 ($\mu\text{S}/\text{cm}^2$)/S	0.2381
Games Not Chosen	-0.0075 ($\mu\text{S}/\text{cm}^2$)/S	0.1525

eliminated, in order to eliminate the effect of the initial orienting response on the general trend of each graph.

Figure 4.25 demonstrates that Mean SCL values for the games that were later chosen in Stage 2 showed an overall increase over time, while SCL values for the games that were not chosen showed an overall decreased over time. Table 4.7 displays the slope and correlation coefficients (r^2 . values) for these lines.

As evidenced by the positive slope of the best-fit line in Table4.7, Mean SCL increased over time in Stage 1 for the games that were later chosen in Stage 2. Conversely, as evidenced by the negative slope of the best-fit line, Skin Conductance Levels, on average, decreased over time in Stage 1 for the games that were not chosen later. This finding supports that of Experiment 4.1, where it was suggested that gradual and sustained increases in SCL over time are indicative of participants enjoyment of a game, while gradual and sustained decreases in SCL over time are indicative that participants do not enjoy that game.

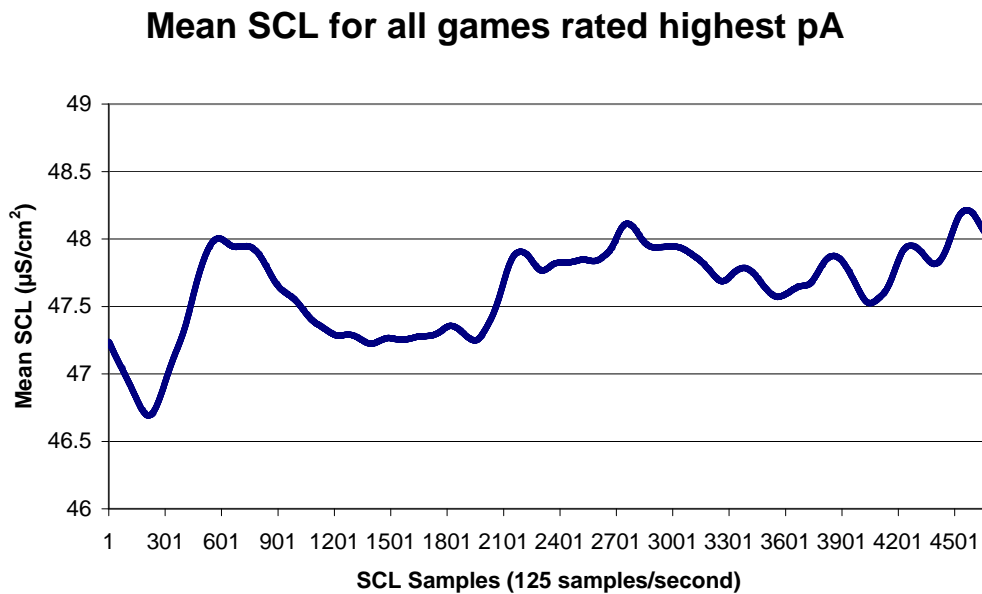


Figure 4.26: Mean SCL recordings for all games in Stage 1 that each participant rated highest for Positive Affect (n=9)

Results of Experiment 4.1 suggested that gradual increases in SCL over time typically accompanied games that were rated as enjoyable using the 12-item self-report measure employed in the current research programme. However, Experiment 4.2 demonstrated how this questionnaire is, in fact, a poor predictor of game preference. Experiment 4.3 subsequently demonstrated that gradual increases in SCL over time typically accompanied games that were later chosen by participants (i.e., the game that was most reinforcing). Thus, the relationship between these dependent measures seems to be unclear. It appears that the signature gradual increase in SCL over time during a game may be indicative of both subjective ratings and observed behaviour.

In order to examine the above assumption, a similar methodology to that used above to examine the relationship of SCL and game choice was adopted to examine the relationship of SCL and ratings of Positive Affect. Specifically, SCL data from the game that each participant rated as highest in Positive Affect in Stage 1 was identified. This data was then compared with the SCL data from all five other games, in order to evaluate whether any distinctive patterns of change in SCL were observed. This analysis was carried out and the results are presented in figures 4.26 and 4.27. Figure 4.26 presents mean SCL data recorded across all games that each participant rated highest in Positive Affect. Figure 4.27 presents mean SCL data recorded across all other games. Neither of these graphs represent data from any one particular game, as different participants rated different games as highest in Positive Affect.

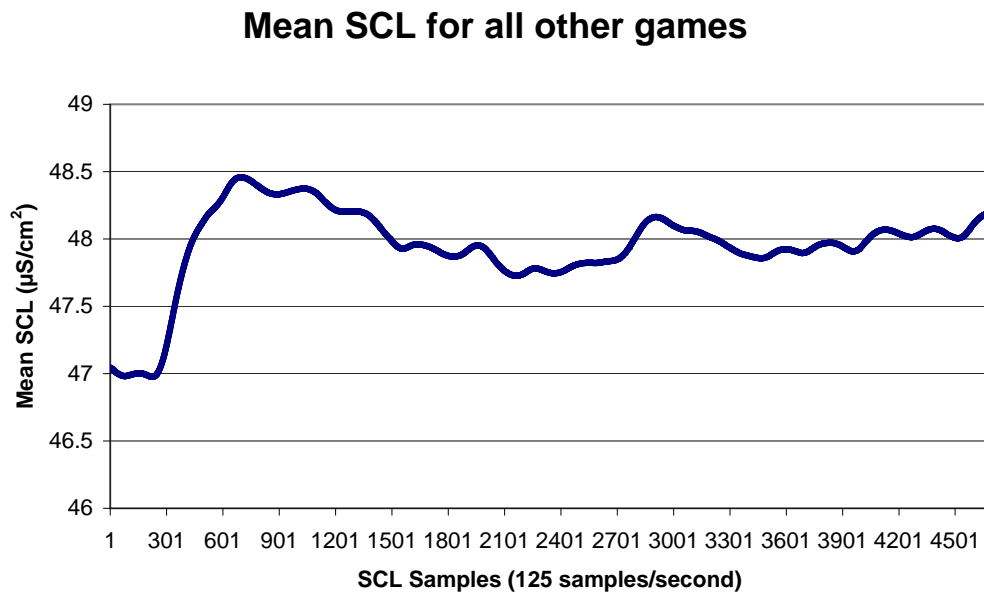


Figure 4.27: Mean SCL recordings for all games in Stage 1 other than those that each participant rated highest for Positive Affect ($n=9$)

Figure 4.26 presents Mean SCL from the game in Stage 1 that each participant rated as highest in Positive Affect. Mean SCL increased sharply over the first five seconds or so, before decreasing sharply, followed by a gradual increase towards the end of the session.

Figure 4.27 presents Mean SCL recordings for all games in Stage 1 other than those that each participant rated highest for Positive Affect. Mean SCL increased sharply over approximately the first five seconds, before decreasing gradually over a period. This was followed by a period where SCL levels displayed a gradual increase until the end of the session.

Data represented by Figures 4.26 and 4.27 appear to be quite different. Specifically, there appears to be a gradual increase in SCL levels across time in the games that each participant rated highest for Positive Affect. Games that were not rated by participants as highest in Positive Affect displayed a large initial increase, followed by a gradual decrease, and then a period where SCL appears to increase gradually towards the end of the session. This finding somewhat supports that of Experiment 4.1, where it was proposed that any gradual increase or decrease in SCL may be indicative of participants subjective ratings of a game.

In order to further illustrate the difference in SCL data between games that were rated by each participant as highest for Positive Affect, and those that were not, best-fit lines were created for the graphs presented in figures 4.26 and 4.27 and the slope and correlation coefficients for these lines were calculated and presented in figure 4.28 and table 4.8. As explained in the results and discussion

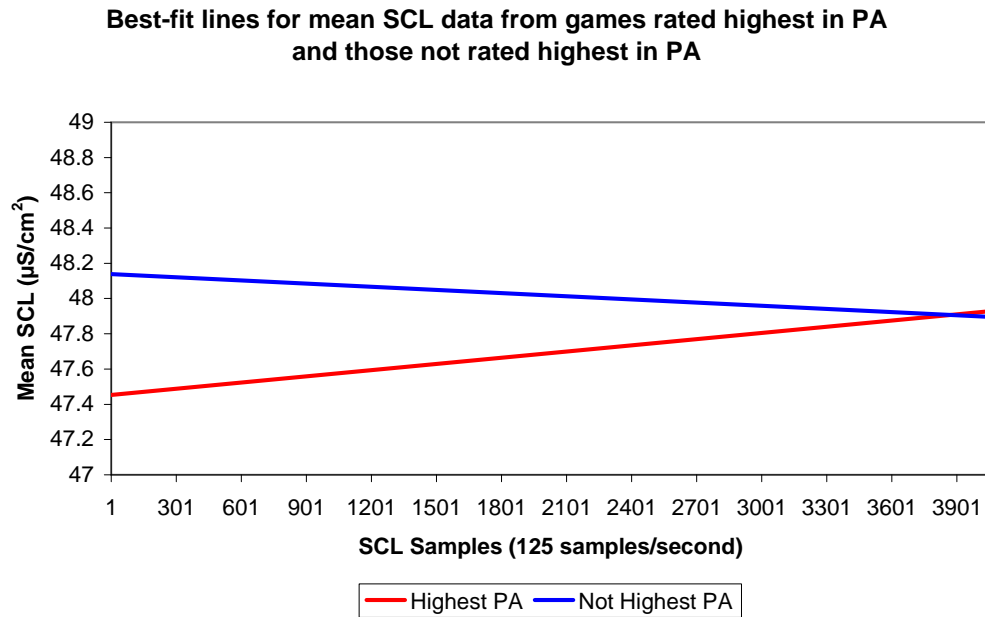


Figure 4.28: Best-fit line for mean SCL recordings for all games in Stage 1 that each participant rated highest for Positive Affect, and from all other games (n=9)

Table 4.8: Slope and correlation coefficients (r^2 . values) for all games in Stage 1 that each participant rated highest for Positive Affect, and from all other games (n=9)

	Slope ($\mu\text{S}/\text{cm}^2$)/S	r^2 . value
Highest Positive Affect	0.0125	0.2611
Not Highest Positive Affect	-0.0075	0.1483

of Experiment 4.1 above, the first five seconds of all EDA recordings were eliminated, in order to eliminate the effect of the initial orienting response on the general trend of each graph.

Figure 4.28 demonstrates that Mean SCL values for the games that each participant rated highest for Positive Affect showed an overall increase over time. SCL values for the games that each participant did not rate as highest for Positive Affect showed an overall decreased over time. Table 4.8 displays the slope and correlation coefficients (r^2 . values) for these lines.

As evidenced by the positive slope of the best-fit line in table 4.8, Mean SCL increased over time in Stage 1 for the games that were rated highest for Positive Affect. Conversely, as evidenced by the negative slope of the best-fit line, Skin Conductance Levels, on average, decreased over time in Stage 1 for the games that were not rated as highest in Positive Affect. This finding supports that of Experiment 4.1, where it was suggested that gradual and sustained increases and decreases in SCL over time are indicative of participants subjective ratings of a game.

Experiment 4.3 found that preference, as measured by free-operant choice, was spread quite

evenly across five of the six games. The patterns of Correct Responses and both Positive and Negative Affect observed in Experiment 4.3 are very similar to those reported in previous experiments. Specifically, number of Correct Responses and ratings of Positive Affect are higher, while ratings of Negative Affect are lower, for the simpler games in comparison to the more complex games. Similarly to the results of Experiment 4.2, correlations between self-report ratings and overt behaviour were low, suggesting that the 12 item scale employed in the current study was not a useful tool for evaluating participants' enjoyment of games.

An analysis of SCL data identified different SCL trajectories across the course of games that participants showed preference for in comparison to those which they did not. Thus, it appears possible to discriminate games which participants enjoyed (i.e., the games that were most reinforcing) from those which they didn't, through an analysis of the SCL data recorded during the playing of those games. Interestingly, using an SCL analysis, it also appears possible to discriminate which games participants will later rate as high in Positive Affect on the 12-item self-report DRM questionnaire, from those which they will not rate as high in Positive Affect. Indeed, the signature pattern of a gradual increase in SCL across time appears to predict that a game will both be rated as enjoyable on self-reports and preferred when a free-operant measure is employed. Conversely, a gradual decrease in SCL across time appears to predict that a game will not be rated as enjoyable on self-reports, nor preferred when a free-operant measure is employed. Thus, a curious relationship appears to exist, where SCL predicts both game choice and subjective ratings, but where ratings do not predict game choice.

4.3.3 General Discussion

Experiment 4.1 demonstrated, using a functional-analytical methodology, that it may be possible to discriminate whether a player is enjoying a particular game from an ongoing analysis of that game players' Skin Conductance Levels. Specifically, it was found that calculating mean SCLs for whole blocks of data and comparing these means across games was of little use in the context of investigating enjoyment in games. Such a methodology did allow for a post-hoc comparative analysis of different games, but did not produce any valuable information regarding enjoyment for any individual game examined in isolation. In summary, mean SCL analyses reveal only which games produce the most and least arousal across the entire game block. This, in turn, reveals little or nothing about which games are comparatively or absolutely enjoyable.

As an alternative to examining mean SCLs, skin conductance was instead tracked over time during game sessions and compared to participants' subjective ratings of those games. Combined

SCL trajectory graphs for all participants were generated for each game, and a best-fit line was generated for each graph and the slope and correlation coefficient of this best-fit line were calculated. This methodology provided the opportunity to identify any characteristic changes in SCL that may have correlated with enjoyable games, irrespective of absolute skin conductance levels. It emerged that the games in which increases in SCL across time were observed, as indicated by a positive slope of the best-fit line, were also the games for which the highest mean Positive Affect ratings and the lowest mean Negative Affect ratings were reported. The games in which decreases in SCL across time were observed, as indicated by a negative slope of the best-fit line, were also the levels for which the lowest mean Positive Affect ratings and the highest mean Negative Affect ratings were reported. It appeared, at least in the preparation employed in Experiment 4.1, that when an increase in SCL across time during a game was observed, that game was later reported as enjoyable. Conversely, it appeared that when a decrease in SCL across time during a game was observed, that game was not reported as enjoyable. This effect was clearly visible at the group level, but also somewhat apparent at the individual level (see Appendix 2), despite the widely recognized noisiness of SCL data when examined on a single subject basis. Specifically, any behavioral paradigm that employs physiological measures instead of “binary” overt operant responses, such as computer keyboard presses, necessarily suffers from the problem of increased noisiness in data (see Dymond, Roche, Forsyth, Whelan, & Rhoden 2007, p. 16; DeHouwer et al. 2005). The use of psycho-physiological measures was important in the current study because a continuous variable measure was needed against which to assess the utility of the DRM measure (i.e., a continuous variable measure).

Experiment 4.1 also provided an alternative definition of complexity than that provided in previous chapters, and examined games that varied in terms of this new definition of complexity. This experiment found that participants gained significantly higher scores on simpler games and rated these games as significantly higher in Positive Affect, and lower in Negative Affect than more difficult games. These findings support those of experiments reported in previous chapters, in suggesting that game players prefer simple games in which they can easily gain high scores. This suggestion has implications for the computer games industry and does not, at least at face value, sit well with some of the literature on computer games, which often identifies ‘challenge’ as a key determinant in game players’ enjoyment of a game (i.e., Gingold, 2005; Malone, 1982; Morlock, Yando, & Nigolean, 1985; Vorderer, Hartmann, & Klimmt, 2003). However, two factors were pointed out that may have influenced this finding; the fact that the questionnaire employed had not been standardized for investigating enjoyment in games, and the fact that each game was presented only once in each experiment. The latter issue may provide an explanation as to why such simple games

were enjoyed by participants. That is, by presenting these games only once the experiment may have confounded simplicity with novelty. In effect, the novelty of even the simplest games may in itself constitute a form of “challenge” to the novice player. These two issues were addressed in Experiment 4.2.

Experiment 4.2 aimed to address uncertainty over the validity of using the 12-item DRM questionnaire in the context of measuring enjoyment in computer games. Specifically, a procedure was developed in which all six games from Experiment 4.1 were presented and participants were asked to complete the 12-item questionnaire after each of these games. Participants were then given free operant choice over which game they wished to play again. As the choice participants made necessarily defined that participants’ favourite (i.e., most reinforcing) game, ratings of Positive and Negative Affect were correlated with this choice, in order to evaluate the validity of the questionnaire. It was found that ratings of Positive Affect predicted actual game preference in approximately 33% of cases. Ratings of Negative Affect produced similar correlations with actual game preference. Thus, Experiment 4.2 suggested that the 12-item scale employed in the majority of studies in the current thesis was not a useful tool for evaluating participants’ enjoyment of games.

Experiment 4.2 also aimed to address the uncertainty over any procedural factors that may have lead participants to show preference towards the simpler games over the more complex games. Specifically, the procedure employed in Experiment 4.1, and a number of experiments in previous chapters, involved the presentation of each game under investigation only once to each participant. It was suggested, due to the novelty of all games presented, that participants may have shown preference for the simpler games, because these games offered the opportunity for participants to gain high scores while the more complex games did not. It was hypothesized that a shift in preference from simple to complex games may have been observed upon repeated presentations of the same games, as participants gained expertise at the more complex games. The free-operant measure employed in Experiment 4.2 allowed the experimenter to observe that upon repeated presentations, preferences did in fact shift from simpler games to more complex games. In addition, the self-report measure also found a general trend, in which the simpler games were rated as less positive and more negative across successive presentations, while the more complex games were rated as more positive and less negative across successive presentations.

These findings qualify the findings of Experiment 4.1, and studies reported in previous chapters, by suggesting that simpler games may only be preferred games by players who are inexperienced with that game. In effect, preference for games may be a dynamic shifting dependent variable that is a function, not only of complexity and speed, but also of experience. This particular conclusion sits

more comfortably with the idea that games need to present a “challenge” to players in order to be enjoyable. That is, it may be that the type of game that constitutes a challenge changes across time in tandem with the changing skill levels of the increasingly experienced player. To that extent, the procedure employed in Experiment 4.2 would appear to provide a good starting point for a more in-depth analysis of the relationship between these variables in studies employing a variety of analog and real-world games.

Experiment 4.1 found that the particular pattern of change in SCL over time may discriminate games that participants rate as enjoyable from those they do not. However, Experiment 4.2 found that the subjective ratings used to assess those enjoyment levels were poor predictors of participants’ actual preferred (i.e., chosen) games. Thus, Experiment 4.3 was conducted in order to assess whether the particular patterns of change in SCL over time identified in Experiment 4.1 could be used to predict participants’ actual game preference, using a procedure similar to that employed in Experiment 4.2. Results from Experiment 4.3 suggested that SCL data recorded during games that were later chosen by participants typically displayed a characteristic gradual increase in Mean SCL, as expected, given the findings of Experiment 4.1. Similarly, those games that were not later chosen displayed a typically gradual decrease in SCL over the course of that game when first played. As noted in the discussion of Experiment 4.1, this effect was clearly visible at the group level, but also somewhat apparent at the individual level (see Appendix 2). Thus, it appears that outcomes on both subjective ratings and behavioural measures of preference can be predicted from the particular patterns of change in SCL recorded during the playing of a game.

A number of general issues have been identified over the course of the current chapter. Firstly, in Experiment 4.2, it emerged that participants’ accuracy and fluency in responding on the more complex games approached that observed in the simpler games, across repeated presentations of all six games. Thus, it is apparent that skill level rose rapidly across successive presentations of the games presented. It may be deduced from this that “challenge” is decreasing across successive presentations. Of course “challenge” is a non technical term and at this stage represents a hypothetical concept. Nevertheless while the term cannot be used here in an empirical way, it can serve as a guiding concept that might inform what types of experiments need to be conducted to pursue these issues in future.

More interestingly, however, the experimental preparation employed in Experiment 4.2 allowed an investigation of precisely what we mean by “challenge” in a technical sense. Specifically, it may be possible to conceive of “challenge” in terms of a skill/score ratio. Indeed, it may be possible to validate any technical definition of the term using the free-operant choice and EDA methodologies

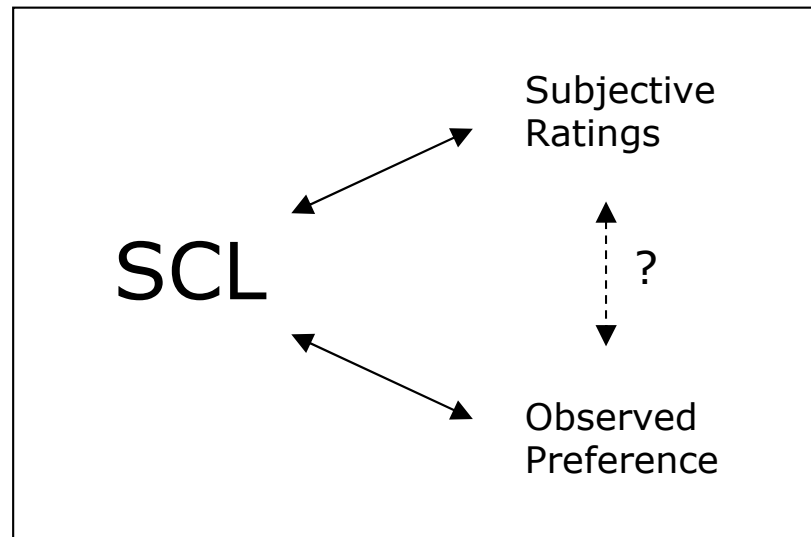


Figure 4.29: Illustration of the predictive relationships between the dependent measures employed in Experiments 4.1-4.3

employed in the current chapter. Pursuing this line of enquiry would represent a technical and conceptual challenge that is beyond the scope of the current research programme. Nevertheless, the fact that such a clear line of investigation is now becoming apparent, and the fact that the current findings are presenting potential technical definitions of previously non-technical terms suggest that the current investigations have been scientifically fruitful.

Secondly, one important outcome of the current research is that SCL trajectories appear to be an excellent index of both future game choices and ratings. However, it appears that subjective ratings are not a useful index of game choice. Figure 4.29 presents an illustration of the relationship between these three dependent measures observed across Experiments 4.1-4.3. The solid lines represent empirically demonstrated predictive relationships. The dashed line represents relationships which have been demonstrated as non-predictive. For example, it is possible to predict the specific pattern of change in SCL recorded from a player given their subjective ratings, and vice versa. Similarly, it is possible to predict the specific pattern of change in SCL given game choices, and *vice versa*. It is not possible to predict a participant's game choice from subjective ratings, and *vice versa*.

It is perhaps not surprising that verbal ratings have fared so poorly as predictors of overt behaviour (i.e., game choice) in the current experiments. From a behavioural perspective, verbal reports are governed by contingencies that are established by the question asked rather than private internal states (e.g., enjoyment) to which the question refers. These contingencies include things like experimental demand and social desirability effects. As a result, it is difficult to know in any one

instance which features of a game, internal states and/or contingencies established by the form of a question (e.g., the way in which it is phrased) control verbal reports. In contrast, EDA levels are controlled by contingencies beyond a participant's awareness. As an index of enjoyment, EDA is unhampered by social or verbal processes. Thus, as it is a more direct measure of the correlates of enjoyable game play, it should function as a more reliable measure of future choice than a verbal report.

It is important to understand that linguistic and conceptual debates regarding the true meaning (i.e., function) of the subjective ratings provided by participants are not relevant to the simple matter of predicting and influencing behavior. The important issue, for both the experimental psychologist and those in gaming industry, is to be able to predict future game and game level choices of individuals and groups, rather than to secure positive ratings as an end in themselves. In effect, given the current findings, researchers would do better to employ the current functional-analytic psycho-physiological approach to game choice than to use verbal reports for such purposes.

Given the foregoing, it would appear that the SCL trajectory observed during game play can be considered a core variable or, "factor," to which other responses to game play are related. This sits well with psychophysiological literature which finds EDA to be the best index of physiological arousal in general and therefore an excellent measure of emotional activity broadly defined (Dawson et al., 2000; Edelberg, 1972). That is, EDA correlates highly with other physiological measures of a variety of emotional states even where these measures do not correlate well with each other.

It would also appear that the verbal reports recorded using the 12-item DRM questionnaire in the current studies are not predictive of game choice, even if they reflect characteristic patterns of arousal associated with enjoyment. Notably, the value of any measure of game enjoyment in research of this kind must be assessed on its ability predict game choice and ultimately purchase. The ability to predict arousal levels is incidental and of little use to further research into game playing. As such, the 12-item self-report questionnaire would appear to be in question as a measure, as are all verbal reports from a behavioural perspective (but see Cabello, Luciano, Gomez, & Barnes-Holmes, 2004; Cabello & O'Hora, 2002; Hayes, 1986).

Interestingly, it has been argued that EDA is not an appropriate measure for the purposes of investigating computer game playing. For example, Sykes and Brown (2003) pointed out that increased perspiration and muscle tension necessarily accompany fast paced arcade-style games requiring, "quick fingered dexterity" (p.732). They argue that the electrical resistance of the skin will change if a player tightens a muscle or perspires heavily, thus interfering with SCL

measurement. Indeed, these issues may render an SCL analysis ineffectual if the techniques employed involve comparing raw SCL values at pre-defined intervals, or comparing mean SCL values recorded during game phases (as seen in Experiment 4.1), rather than examining patterns of SCL change across time. Any technique that involves reducing the dynamically and constantly fluctuating pattern of EDA data to a few representative data points fails to capture the functional nature of any piece of behaviour by removing the defining temporal properties of behavior itself.

The technique employed in the current study involves calculating the trajectory of SCL data over time. Importantly, this does not involve merely sampling data at the beginning and end of each session and drawing a line between these two points. The technique involves identifying any characteristic patterns of change in constant SCL data across time. This emphasis on moment-to-moment change, rather than absolute levels allowed the experimenter to identify important patterns in the data which may otherwise have been missed. Most importantly, it was noticed that it was the gradual increase in SCL across time, rather than the difference in SCLs in the earlier and later phases, which characterises the signature “enjoyment” graph.

The methodology employed in the current study appears to circumvent the two problems with using SCL in game research that were pointed out by Sykes and Brown. Firstly, while large momentary changes in SCL may be observed upon participants’ tensing of a muscle, the particular technique employed in the current study was not concerned with phasic (i.e., brief) changes in SCL, but with the overall trend in SCL across an extended period of time. Thus, over a period of 20-30 seconds, any momentary increases and decreases in SCL contribute little to the overall trajectory of the graph.

Secondly, Sykes and Brown (2003) point out that increased perspiration during game play due to thermoregulatory rather than psychological responding may affect the reliability of the SCL as a measure of enjoyment in games. However, activation of sweat glands and perspiration are key components of the ANS response to a stressor, and actually constitute what is measured by EDA. In effect, sustained activation of the sweat glands, over the period of game-play, will necessarily correlate with increased perspiration and increased SCL. Moreover, thermoregulatory sweating on the palmar surfaces of the human body by both eccrine and apocrine glands rarely occurs (Fowles, 1986; Jakubovic & Ackerman, 1985). Thus, any sweat produced by palmar sweat glands is unlikely to be a result of thermoregulatory function and likely to be a result of psychological processes. The particular technique developed in the current chapter for analysing SCL data appears to have yielded potentially important findings despite the concerns raised by Sykes and Brown over the suitability for using EDA in the analysis of game enjoyment.

Of potentially great importance, Experiments 4.1 and 4.3 provided specific slope-of-graph gradients that may turn out to be useful as an absolute coefficient for indicating future game choices, if not enjoyment. In other words, it is the slope of the arousal increase across time, rather than absolute levels of arousal that provides the most interesting data for the current purposes. Future research with a wide variety of different games may help to further determine whether the SCL slopes observed here for enjoyable and chosen games are applicable to a wider variety of enjoyed and chosen computer games.

Another issue that would appear to merit further investigation relates to the cycle of arousal that would be observed if subjects were exposed to the game over longer periods of time. Presumably, the SCL levels of a game that appears to sustain attention must eventually drop. The point at which this happens may also provide information about the functional stimulus properties of the game for individual subjects (i.e., how long the game can hold their attention). Despite the foregoing outstanding empirical questions, the functional analytic approach employed in the current chapter proved to be useful in suggesting a quantitative technique that may be of interest to those in the gaming industry or researching the psychology of game playing.

Of potential interest to a wider audience, it appears that, as expected, computer games could be conceived of as tasks requiring sustained attention. A large body of research exists that identifies a gradual increase in physiological arousal as an accompaniment of sustained attention (see Bohlin, 1976; Grim, 1967; Oken, Salinsky, & Elsas, 2006; Smallwood, Davies, Heim, Finnigan, Sudberry, O'Connor, & Obonsawin, 2004). Precisely this pattern of arousal was observed in the current studies during games that participants both chose to play later and rated as enjoyable. Thus, physiological data suggests that the games that participants enjoyed were the games which sustained participant's attention. These findings provide a possible conceptual bridge between the current behavioural research on enjoyment in computer games, and a standard and well-researched concept in cognitive psychology. Indeed, both fields may benefit from cross-fertilisation of these obviously parallel ideas; that sustained attention and the potency of reinforcing experiences (i.e., enjoyment) may be closely related.

All of the foregoing issues point to the fact that many more conceptual and empirical analyses remain to be done to further explore the issues investigated in this thesis. Nevertheless, the results of Experiments 4.1-4.3 in the current chapter would appear to be of immediate practical importance. Specifically, it may be possible to use real-time input from an EDA device in order to maintain enjoyment/challenge for game players of varying skill levels, without needing players to make explicit game level choices. In effect, the game itself could use EDA data to control game complexity to suit

individual players' preference and ability. Some researchers (i.e., Gilleade & Dix, 2004) have pointed out the value of such a finding. In the words of Gilleade and Dix; "Preferably videogames should be capable of dynamically changing their design (i.e. adaptive) in light of the players' ongoing interactions with the videogame" (2004, p.228). While Gilleade and Dix recognized the need for adaptable games, they rejected the use of physiological responses as input, citing the concerns raised by Sykes and Brown (2003). However, as outlined above, these potential limitations may now be surpassed using the functional analytic approach employed in the current research.

It must be pointed out that considerable research has already been conducted on using EDA input to control game-play in games that use biofeedback to let players control their psychophysiological response (i.e., Bersak, McDarby, Augenblick, McDarby, McDonnell, McDonald, & Karkun, 2001; Parente & Parente, 2006; Sakurazawa, Yoshida, Munekata, Omi, Takeshima, Koto, Gentsu, Kimura, Kawamura, Miyamoto, Arima, Mori, Sekiya, Furukawa, Hashimoto, & Numata, 2003). Indeed, there is at least one game that uses this technology, *The Journey to Wild Devine*, (Wild Devine) currently commercially available. However, in these biofeedback games the psychophysiological input is not used as additional information that allows the game to adapt to the players state of arousal. Rather, these games concentrate on helping players to gain control of their psychophysiological responses for health purposes. The graphics displayed on-screen are merely graphical representations of the players' psychophysiological responses. Game advancement simply involves gaining finer control over physiological responding. Thus, as evidenced by the existence of commercially available biofeedback games, the technology to build games that are adaptive to individual participant's preference and ability already exists. However, it appears that researchers have not yet been able to create games that adapt to individual user's arousal levels. This would appear to be due to a lack of knowledge regarding the rigour of experimental analysis of human behaviour research and the use of a functional-analytic approach to analyzing psychophysiological data.

The findings of the current chapter are of immediate practical importance to computer games researchers and designers. Experiments conducted in the current chapter demonstrated that SCL can be used to measure and predict participants' preference for particular games. A signature SCL trend has been identified that appears to describe game enjoyment and predict game choice. Additionally, a distinct SCL trend was identified that describes dislike for a game and predicts that the game will not be chosen to be played. Using this information, it now seems possible to build a game that is adaptive to participants' game preference and ability levels. SCL technology is cheap, can be attached to the body non-intrusively, can be sold with a game, and interfaced with the game

easily, as demonstrated by the success of *The Journey to Wild Devine*. Thus, games could be created that are sold with an optional SCL add-on that tracks SCL during game play, and searches for the two SCL signatures identified in the current chapter. This information could then serve as an input to the games, so that complexity levels are maintained so long as the signature enjoyment pattern of gradually increasing SCL is observed. Complexity could then be increased if the SCL trend stabilises, or shows a decrease across time during game-play. If the signature can not be re-established by increasing complexity, complexity could be decreased until the signature pattern is re-established. Thus, it appears that the current research has laid the technical and conceptual groundwork for a whole new set of research opportunities within the gaming and human-computer interaction fields that may well lead to a new and exciting generation of adaptive computer games, currently just beyond the horizon of game developers.

Chapter 5

General Discussion

5.1 Chapter 2 - Experiments 2.1 and 2.2

In Chapter 2, four experiments were presented that represent the first step in a systematic behavioural investigation of on-line computer gaming. These studies demonstrated the processes of stimulus generalization and derived relational responding in the context of a simple computer game, and thus suggest how important features of computer game playing may be understood and analysed in behavioural terms. The four studies presented in Chapter 2 also examined the effect of Network Latency on players' game playing behaviour, including both their objectively measurable success as well as their subjective assessment of game quality.

All participants who completed the Training Level in Experiments 2.1 and 2.2 scored at above chance in the Generalisation phase. This finding demonstrates stimulus generalisation in the context of game playing and it is offered here as a suitable analogue of some low-level skills required by rudimentary games. In addition, all but two participants in Experiments 2.1 and 2.2 who completed the training phase scored above chance levels in the equivalence phase. In order to gain a high score in the latter phase, participants were required to demonstrate a transfer of the appropriate response function through derived equivalence relations. The finding that a large number of participants were able to consistently respond in accordance with the previously learned equivalence relations provides support for the view that game players can show derived relational responding in the context of computer games.

5.1.1 Network Latency

Simulated Network Latency, as variously defined across Experiments 2.1 and 2.2, had a significant effect on level of correct responding as well as on subjective measures of difficulty and frustration. Participants found the Delay Game significantly more difficult and frustrating than the Non-Delay Game, and achieved significantly lower scores when playing the former than the latter. It is important to appreciate that these effects were replicated across two separate experiments, suggesting that these findings are quite reliable and easily approached and analysed using the current derived relations procedures.

Despite the foregoing, simulated Network Latency was not found to influence participants' enjoyment ratings of games. This unexpected outcome could be interpreted in two ways. Firstly, it may be suggested that enjoyment of a game is uncorrelated with success at that game, or the difficulty and frustration experienced. However, this interpretation would appear to contradict a number of research findings (e.g., Malone, 1982; Vorderer, Hartmann, & Klimmt, 2003), which suggest that the level of difficulty of a game and enjoyment experienced with that game are closely related.

Secondly, it could be argued that reliable and valid data is always difficult to obtain using non-standardised subjective rating scales, such as those employed in the current study. It must be remembered, however, that the experiments in Chapter 2 were both novel and exploratory and were intended to pave the way for the development of reliable subjective measures that may be validated against objective empirical information regarding the experience of on-line gaming. In effect, it is precisely by arriving at the limits of the experimental design and dependent measures employed that we can begin to develop the appropriate measures for the analysis of game playing at a quantitative level. Thus, one of the issues arising from the findings of Experiments 2.1 and 2.2 is that more reliable measures of game playing experience were required for the current research.

5.1.2 Relational Complexity

Perhaps the most interesting finding of the study was that neither Experiment 2.1 nor 2.2 found a significant effect for Relational Complexity on game performance or on any of the subjective measures. This finding suggests that increasing complexity, at least as conceived in the current study, does not alter the experience of gaming by increasing levels of enjoyment, difficulty or frustration. However, these findings may also suggest that complexity was not successfully manipulated in Experiments 2.1 and 2.2. It could be argued that the levels of relational responding examined in these experiments were relatively low compared to those required in a large number of

modern games. In effect, it might not be that participants enjoy less relationally complex games across the board, but appear to do so only when the levels under examination are low and relatively non-stimulating.

5.2 Experiments 2.3 and 2.4

Experiments 2.3 and 2.4 were conducted in order to address three issues which arose in the analysis of Experiments 2.1 and 2.2. Firstly, Experiments 2.3 and 2.4 aimed to extend the analysis of Relational Complexity conducted in Experiments 2.1 and 2.2 by increasing the complexity of the relational activities required to play the games. This was carried out through manipulating the size of the stimulus classes in both experiments. Additionally, the stimulus equivalence testing phase was removed from the procedure of Experiment 4 in an attempt to clarify whether administering an equivalence test before game play was a confounding factor in the impact of Relational Complexity on game enjoyment.

Secondly, Experiments 2.3 and 2.4 were designed to extend the analysis of Network Latency carried out in Experiments 2.1 and 2.2, by extending the length of delays in the games. This strategy was adopted due to findings from the field of engineering (i.e., Delaney, Meenaghan, Ward & McLoone, 2004; Vaghi, Greenhalgh, & Benford, 1999;), which suggest that incrementally increasing the level of delay in a game leads to a progressively less ‘playable’ game.

Thirdly, it was unclear in Experiments 2.1 and 2.2 whether the expected effects were occluded to some extent by the non-standardised rating scales employed. Thus, an alternative rating scale was employed in Experiments 2.3 and 2.4 that was developed in order to measure subjective experience of recent events. Due to the fact that extensive research had been carried out on this instrument (see Kahneman et al., 2004), it was expected to represent a more reliable subjective measure of enjoyment than that previously employed.

As in Experiments 2.1 and 2.2, participants in both Experiments 2.3 and 2.4 demonstrated robust derived relational responding in a game playing environment. Indeed, participants in Experiment 2.3 had relatively little difficulty in reaching a 90% correct response criterion on a game involving derived three-node equivalence relations. This behaviour, therefore, may serve as a model of the types of complex behaviours observed in some online computer games.

5.2.1 Relational Complexity

Relational Complexity significantly affected scores in Experiment 2.3. Specifically, participants attained lower total correct responses in the later levels of the games than in the earlier levels. These findings confirm that through manipulating the level of derived relational responding required in a game, the level of difficulty or challenge experienced by the players can be increased. As such, this finding supports the derived relations approach to understanding users' experience of online games and confirms that difficulty of a game can be manipulated experimentally in a quantifiable and functionally understood way.

It was argued that experimental control may have been compromised in Experiment 2.3 by the administration of an equivalence test prior to the game stage. Such a test requires participants to derive both the one-node and three-node relations in advance of the game. Therefore, the procedure employed in Experiment 2.3 may have negated the need for participants to derive the appropriate relations during the game play itself. As such, a stimulus equivalence test was not administered before game play in Experiment 2.4.

In Experiment 2.4, no significant difference was found between scores on the earlier levels of the game and those on the later levels. This finding may be explained by the relatively high performance demands placed on participants in Experiment 2.4 compared to Experiment 2.3. This high demand was generated by the absence of a stimulus equivalence testing phase in the procedure of Experiment 2.4. Essentially, participants in Experiment 2.4 were required to engage in a complex problem solving task under demanding response time constraints. This may have led to a floor effect on scores in both levels 2 and 3 in Experiment 2.4.

Relational Complexity significantly affected Positive Affect ratings in both Experiments 2.3 and 2.4. Specifically, in both experiments, there was generally a trend of lower Positive Affect ratings for the more complex levels, suggesting that participants preferred the relationally less complex levels over the more complex levels. The finding that participants prefer to engage in less challenging games may be of considerable interest to the computer games industry. Specifically, these findings challenge the conventional wisdom in the engineering literature that suggests a link between higher levels of challenge and enjoyment in games. It must be pointed out, however, that challenge, or complexity, were defined here only in relational terms in the current study. Nevertheless, this in itself represents an important contribution. More specifically, while people on the whole may prefer more challenging to less challenging games, they may not necessarily want challenge to be presented in terms of Relational Complexity. It was suggested that other forms of complexity, such as time constraints, precise motor skills, the Number of Stimuli involved, and so on, may make for an

enjoyable game before Relational Complexity.

Of course, it could be argued that even the levels of relational responding examined in all experiments in Chapter 2 were relatively low compared to those required in a large number of modern games. This is borne out by the high number of correct responses observed in all games at all levels. In effect, as mentioned above in the discussion of Experiments 2.1 and 2.2, it might not be that participants enjoy less relationally complex games across the board, but appear to do so only when the levels under examination are low and relatively non-stimulating. It was argued that a game in which participants demonstrated clearer evidence of a struggle to produce the correct answer under time constraints may produce a different pattern of results.

It is important to note that there does not appear to be a link between the scores that participants achieved while playing the games in both Experiments 2.3 and 2.4 and their subjective ratings of those games for Positive and Negative Affect. Thus, it appears that the scores participants achieve in a game do not necessarily determine their enjoyment of that game. At least in the current preparation, it appears that the two dependent variables which represent success at a game and enjoyment of that game are independent. This, in itself, is an important empirical finding for the computer games industry and the field of psychology more generally.

5.2.2 Network Latency

Experiments 2.3 and 2.4 were conducted in order to extend the analysis of Network Latency conducted in Experiments 2.1 and 2.2. Specifically, participants were presented with a game containing either short 0.5s delays or long 1s delays. No significant difference was found between these groups, on any measure employed in the study. This finding would suggest that there is no functional distinction between the 0.5s and 1s delays modelled in those experiments. However, these results must be viewed with caution, as there was no control or 'No Delay' condition. Therefore, these results do not suggest that Network Latency has no effect on participants' performances or enjoyment of games more generally. In fact, in Experiments 2.1 and 2.2, ratings of difficulty and frustration were significantly higher for participants who played a game containing simulated network delays than for those who played a game containing No Delays. In addition, participants who played the game containing delays achieved significantly lower scores on that game. Taken together, the results of these four studies may suggest that while the presence of delay in a game is destructive to the game playing experience, increasing the length of that delay beyond 0.5s does not have any further negative effects.

It must be noted that the method in which Network Latency was modelled in Chapter 2 may

have lacked ecological validity. Specifically, Network Latency was modelled as a fixed interval of either 0.5 seconds or 1 second. In practice, Network Latency is rarely, if ever, predictable, and typically oscillates erratically during game play. It has been suggested that this oscillation in Network Latency, known as jitter, is much more destructive to the game playing experience than fixed delays (Delaney, Meenaghan, Ward & Mc Loone, 2004), such as those modelled in Chapter 2. Thus, an attempt to simulate the effect of jitter as part of any further investigations of Network Latency was necessary. Chapter 3 addressed this and several of the foregoing issues.

5.3 Chapter 3 - Experiments 3.1 and 3.2

The four experiments presented in Chapter 2 were designed to analyse the functional relationships between the variables of Relational Complexity and Network Latency, and to assess the effects of these variables on a number of measures. The analysis has revealed that the effects of complexity may be observed in only a narrow range of game types that are neither very high nor low in complexity, and that network delay appears to interact in a non-linear fashion with complexity. Indeed, it appears that both floor and ceiling effects may have been in operation in the Chapter 2 experiments. As a result, Experiments 3.1 and 3.2 were conducted in order to gain greater control over the variables of Relational Complexity and Network Latency than achieved in previous studies. The two variables of interest were examined separately in Experiments 3.1 and 3.2 respectively, so that the effects of each could be identified independently.

5.3.1 Experiment 3.1 - Relational Complexity

Experiment 3.1 aimed to model a game involving derived relations that presented different levels of difficulty to participants, and measured the effect that these different levels of difficulty had on participants' performance and their subjective ratings of enjoyment. In Experiment 3.1, Relational Complexity was examined in terms of derived SAME and OPPOSITE relations. This represented a more complex form of derived relational responding than that employed in previous studies. Importantly, it was argued that the use of SAME and OPPOSITE relations provided a functional definition of level of complexity in a game (see introduction, Chapter 3). Specifically, an OPPOSITE relation requires a more complex form of relational responding than a SAME relation (Barnes & Hampson, 1993; O'Hora et al., 2002; Steele & Hayes, 1991; Whelan & Barnes-Holmes, 2004; Whelan et al., 2006). Thus, a level which requires players to engage in derived opposite responding is more complex than one which requires participants to engage in derived equivalence, or SAME responding.

Experiment 3.1 found that participants produced significantly fewer correct responses on the relationally more complex levels than in the less complex levels, suggesting that those levels did represent significantly different levels of difficulty for participants. In addition, it was noted that the majority of participants produced scores in the SAME and OPPOSITE level that would typically constitute a passing criterion on a relational testing phase. While this level of correct responding was not reached on the most complex level (i.e., the Mixed SAME and OPPOSITE Level), more than half of the participants did produce more correct responses in this level than would be expected by chance. Thus, it does appear that a game involving derived relational responding was modeled successfully in Experiment 3.1.

In addition, there was a trend of higher negative and lower positive subjective ratings in the more complex levels. Indeed, while this pattern was not significant at the 0.05 level, it was linear across game levels and across the majority of participants. The SAME game was rated as more positive and less negative than the OPPOSITE game. The OPPOSITE game, in turn, was rated as more positive and less negative than the mixed SAME and OPPOSITE game. In summary, there was once again a tendency for participants to prefer relationally simpler games to more complex games.

Experiment 3.1 was devised in order to clarify the impact of Relational Complexity on performance and enjoyment of computer gaming, as previous studies were affected by both ceiling effects (i.e., Experiments 2.1 and 2.2) and floor effects (i.e., Experiment 2.4). Of the previous experiments, only Experiment 2.3 represented a successful manipulation of the complexity variable, insofar as scores differed significantly across levels. Results of both subjective and objective measures from that experiment parallel those from Experiment 3.1. In both of these experiments, participants produced lower scores on the relationally more complex levels, while rating those levels as less positive and more negative than the less complex levels. Thus, it appears that the findings of Experiment 3.1 further support the idea that participants may prefer less relationally complex games over more relationally complex games.

It must be noted that the patterns of both objective and subjective data observed were not universal in either Experiment 3.1 or any of the previous experiments. It is possible that the effect of complexity may vary across participants in tandem with levels of other variables, such as experience at game playing, extent of history of derived relational responding in similar contexts, or as yet unknown physiological variables. Under these circumstances, it is difficult to make generalisations for the research literature or the gaming industry as to the general levels of complexity that will produce decrements in performance and game enjoyment for most or all participants. Nevertheless,

taking into account results from the current and previous experiments, it appears that participants prefer to play relationally less complex games.

Once again, the finding that participants prefer to play relationally less complex games was unexpected. However, as suggested previously, the finding does not necessarily indicate that all complexity is not reinforcing. Rather, it may simply be that complexity involving the solving of a cognitive problem (e.g., deriving relations), ‘on the hoof’ during game-play is not reinforcing. While some individuals may choose games involving complex cognitive tasks (i.e., strategy games) over simpler games, it appears that most participants do not respond this way in the current experiments. Of course, some degree of challenge may well be necessary for a game to be experienced as enjoyable. Indeed, this assertion is in line with research which suggests that participants most enjoy games with an appropriate but measured level of ‘challenge’ (i.e., Gingold, 2005; Malone, 1982; Morlock, Yando, & Nigolean, 1985; Vorderer, Hartmann, & Klimmt, 2003). However, no technical or psychological definition for the term ‘challenge’ is available in the literature and that it is unclear whether or not the most enjoyable degrees of ‘challenge’ are relational in nature, or merely spatial and temporal. Nevertheless, it was argued that a conceptually sophisticated theoretical framework and a sufficiently technical research methodology, such as that provided by RFT, should allow us to accurately distinguish the various dimensions of complexity that may be conceived. In so doing, researchers will be in a position to compare the relative effects of various dimensions of complexity on the gaming experience. Chapter 4 was designed to address this very issue.

Importantly, the finding that participants produced a significantly different amount of correct responses across different levels of the game in Experiment 3.1 implies that the game format may provide a novel measure of fluency with derived relations in behavioral research more generally. Specifically, the unique game environment required participants to respond with both accuracy and speed in order to gain a high score. Indeed, such a methodology may be useful for people in applied contexts who are not just interested in acquisition of derived relations but in application of those relational training techniques in real world contexts. For example, the game could represent a user-friendly method of assessing relational skills in a range of children with developmental delays. In such a preparation, the game could be used in place of a standard relational testing phase. Furthermore, the game format could be used as an intervention to improve the skill of deriving relations as suggested by literature on Multiple exemplar training (Berens & Hayes, 2007; Hayes, Barnes-Holmes, & Roche, 2001a; Murphy, Barnes-Holmes, & Barnes-Holmes, 2005; Gómez, López, Martín, Barnes-Holmes, & Barnes-Holmes, 2007). Indeed, it would appear that the current game-based presentation format may actually facilitate high levels of accurate responding which

could be called exceptional by those experienced in the field of equivalence training and testing (see Arntzen & Holth, 1997, 2000; Fields, Adams, Verhave, & Newman, 1990; Fields, Reeve, Rosen, Varelas, & Adams, 1997; Saunders, Saunders, Williams, & Spradlin, 1993; Smeets & Barnes-Holmes, 2005).

5.3.2 Experiment 3.2 - Network Latency

Experiment 3.2 examined the impact of Network Latency on game performance and enjoyment through games that varied from containing No Delays to games which contained delays on all trials. In addition, 'jitter, (i.e., the unpredictable variation in the length of network delays) was modeled in a more ecologically valid method than previously achieved. Participants did not produce significantly more or less correct responses across the three levels in the game. In addition, Network Latency only had a significant effect on one of the four subjective measures employed. However, the single subject analysis revealed that the group design statistics masked some trends in the data, which were not evident for all participants, but which were quite consistent for a large number of participants. Specifically, it appears that while the No Delay level was rated as more positive, higher in competence, less negative and lower in impatience than either the Half Delay or Full Delay levels, these levels were rated as similar to each other across all four subjective measures. These results appear to replicate the combined findings of Experiments 2.1-2.4 in one single experiment, in suggesting that while participants prefer to play games which do not suffer from any Network Latency, a game which suffers from a large amount of latency is not less enjoyable than a game which suffers from a modest amount of latency. This finding may have implications for the games industry, in that it demonstrates the necessity for games designers to eliminate the effects of latency on game-play, as a small amount of latency appears to be just as detrimental as a lot of latency.

Interestingly, the finding that a small amount of latency appears to be just as detrimental as a lot of latency is not consistent with findings from the field of engineering (i.e., Delaney, Meenaghan, Ward, & McLoone, 2004; Hashimoto & Ishibashi, 2006; Vaghi, Greenhalgh, & Benford, 1999), that suggest that incrementally increasing the level of delay in a game leads to a progressively less 'playable' game. However, it must be pointed out that the minimum delay employed in the current studies was 500ms. As demonstrated by Delaney et al. (2004), Hashimoto and Ishibashi (2006) and Vaghi et al. (1999), a 500ms network delay would be considered quite large and disruptive to game-play. Engineering researchers have investigated much shorter delays, to the point where they are not perceptible by users. Thus, it may be the case that incrementally increasing delay correlates with decreasing enjoyment of a game up to a point of approximately 500ms, from where it levels off.

5.3.3 General Comments on Chapters 2 and 3

A number of general findings emerged over the course of research conducted in Chapters 2 and 3. Firstly, the studies conducted in these chapters demonstrate that it is possible to model games using the concept of derived relations. Across the six experiments, participants demonstrated stable responding in accordance with stimulus generalisation, one node and three node stimulus equivalence, and complex stimulus relations in a game playing environment. In addition, the technical definition of game ‘challenge’ in terms of Relational Complexity provides a sound behavioural method of differentiating games of varying levels of complexity. Thus, it appears that derived relations may provide a useful conceptual framework for understanding complex game playing as a behavioural activity. Secondly, it was generally the case that participants produced lower scores on the relationally more complex games and rated those games as less positive and more negative than the less complex games. However, these patterns were not observed universally across all participants and game types, and indeed conflicting patterns were sometimes observed across different experiments. Thirdly, the presence of any Network Latency in a game was found to be detrimental to the game playing experience, in that it affected both the scores obtained in, and subjective ratings of, those games. Increasing the length of delays in games generally did not have a significant effect on either correct responses made or subjective ratings.

5.4 Chapter 4 - Experiments 4.1-4.3

Considering the findings of the foregoing chapters, it was suggested that it may not have been prudent to indefinitely continue pursuing the relationships between the variables of Relational Complexity and Network Latency on computer game experience and enjoyment. The variability in the various performances and outcome measures strongly suggested that the relationships between the variables was highly complex. The experiments of Chapters 2 and 3 served to show how complex a task it would be to arrive at clear statements on the linear and stable relationships between the variables of Network Latency, complexity and gaming enjoyment. However, conclusions regarding the linear or nonlinear nature of the effects of delay and complexity are tentative given the persistent problem of poor subjective measures. As long as subjective measures developed by non-behavioural researchers for non-behavioural purposes were being employed, it would be impossible to pinpoint the source of large variances in measures across participants using these measures. It became clear, therefore, that a significant contribution to the research literature could be made if more attention was paid to the development of reliable and objective measures of enjoyment of game playing.

The three experiments reported in Chapter 4 involved the development of novel measures of gaming enjoyment and complexity that should make future research into the effects of game complexity and Network Latency easier to measure, predict and control. The measurement of enjoyment is especially difficult because of a lack of psychometric tests standardised for this purpose, and because verbal reports obtained from paper and pencil tests are not particularly reliable as predictors of future behaviour and assessing their validity represents a major conceptual challenge (see Cabello & O’Hora, 2002; Critchfield & Epting, 1998; Dymond & Barnes, 1997; Perone, 1988, for further discussion).

Chapter 4 aimed to develop novel operant and psychophysiological measures of enjoyment in order to deal with the two foregoing issues. Traditional behavioural methods, such as evaluating the reinforcing properties of different games, represent an obvious means of assessing game preference and may provide more reliable measures of enjoyment in games. Psychophysiological arousal levels may also represent an operational definition and measure of game enjoyment that may prove more reliable and valid than subjective reports. The experiments reported in Chapter 4 were also developed to more closely consider alternative definitions of complexity. In particular, a simpler definition of complexity with greater face validity, in terms of speed of stimulus presentation and the number of stimuli involved was employed to further investigate the relationship between complexity and game enjoyment.

5.4.1 Experiment 4.1

Experiment 4.1 demonstrated, using a functional-analytical methodology, that it may be possible to discriminate whether a player is enjoying a particular game from an ongoing analysis of that game players’ Skin Conductance Levels (SCL). It appeared, at least in the preparation employed in Experiment 4.1, that when an increase in SCL across time during a game was observed, that game was later reported as enjoyable. Conversely, it appeared that when a decrease in SCL across time during a game was observed, that game was not reported as enjoyable. This effect was clearly visible at the group level, but also somewhat apparent at the individual level, despite the widely recognized noisiness of SCL data when examined on a single subject basis.

Experiment 4.1 also found that participants gained significantly higher scores on simpler games and rated these games as significantly higher in Positive Affect, and lower in Negative Affect than more difficult games. These findings support those of experiments reported in previous chapters, in suggesting that game players prefer simple games in which they can easily gain high scores. However, as with previous experiments, the questionnaire employed had not been standardised for

investigating enjoyment in games and the now well cited problems incurred by these measures may have applied to these results.

5.4.2 Experiment 4.2

Experiment 4.2 aimed to address the uncertainty over the utility of the 12-item DRM questionnaire in the context of measuring enjoyment in computer games. Specifically, a reinforcement value assessment procedure was developed in which all six games from Experiment 4.1 were presented and participants were asked to complete the 12-item questionnaire after each of these games. Participants were then given free operant choice over which game they wished to play again. The choice participants made necessarily defined that participants' favourite (i.e., most reinforcing) game. Ratings of Positive and Negative Affect were correlated with this choice in order to evaluate the validity of the questionnaire. It was found that ratings of Positive Affect predicted actual game preference in approximately 33% of cases. Ratings of Negative Affect produced similar correlations with actual game choice. Thus, Experiment 4.2 suggested that the 12-item scale employed in the majority of studies in the current thesis was not a useful tool for evaluating participants' enjoyment of games.

Experiment 4.2 also aimed to address the uncertainty over any procedural factors that may have lead participants to show preference towards the simpler games over the more complex games. Specifically, the procedure employed in Experiment 4.1, and a number of experiments in previous chapters, involved the presentation of each game under investigation only once to each participant. It was suggested, due to the novelty of all games presented, that participants may have shown preference for the simpler games, because these games offered the opportunity for participants to gain high scores while the more complex games did not. The free operant measure employed in Experiment 4.2 allowed the experimenter to observe that upon repeated presentations, preferences did in fact shift from simpler games to more complex games. These findings qualify the findings of Experiment 4.1, and studies reported in previous chapters, by suggesting that simpler games may only be preferred games by players who are inexperienced with that game. In effect, preference for games may be a dynamic shifting dependent variable that is a function, not only of complexity and speed, but also of experience. This particular conclusion sits more comfortably with the idea that games need to present a "challenge" to players in order to be enjoyable (see Gingold, 2005; Malone, 1982; Morlock, Yando, & Nigolean, 1985; Vorderer, Hartmann, & Klimmt, 2003). That is, it may be that the type of game that constitutes a challenge changes across time in tandem with the changing skill levels of the increasingly experienced player. To that extent, the procedure employed in

Experiment 4.2 would appear to provide a good starting point for a more in-depth analysis of the relationship between these variables in studies employing a variety of analog and real-world games.

5.4.3 Experiment 4.3

Experiment 4.3 was conducted in order to assess whether the particular patterns of change in SCL over time identified in Experiment 4.1 could be used to predict participants' actual game preference, using a procedure similar to that employed in Experiment 4.2. Results from Experiment 4.3 suggested that SCL data recorded during games that were later chosen by participants typically displayed a characteristic gradual increase in Mean SCL, as expected, given the findings of Experiment 4.1. Similarly, those games that were not later chosen displayed a typically gradual decrease in SCL over the course of that game when first played. Thus, it appears that outcomes on both subjective ratings and behavioural measures of preference can be predicted from the particular patterns of change in SCL recorded during the playing of a game.

5.4.4 General Comments on Chapter 4

A number of general issues arose from the experiments in Chapter 4. Firstly, in Experiment 4.2, it emerged that participants' accuracy and fluency in responding on the more complex games approached that observed in the simpler games, across repeated presentations of all six games. It is apparent that skill level rose rapidly across successive presentations of the games presented. It may be deduced from this that "challenge" is decreasing across successive presentations. Importantly, therefore, the experimental preparation employed in Experiment 4.2 allows for an investigation of precisely what is meant by "challenge" in a technical sense. For instance, it was suggested that it may now be possible to conceive of "challenge" in terms of a skill to score ratio. Indeed, it may be possible to validate this or any technical definition of the term "challenge" using the free operant choice and physiological methodologies employed in that chapter.

Secondly, the nature of the relationship between, and the relative utility, of each of the three dependent measures employed in the Chapter 4 was clarified. Specifically, it appeared that SCL trajectories were an excellent index of both future game choices and ratings. However, it also appeared that the verbal reports recorded using the 12-item DRM questionnaire were not predictive of game choice, even if they reflected characteristic patterns of arousal associated with enjoyment. It is perhaps not surprising that verbal ratings fared so poorly as predictors of overt behaviour, as verbal reports are not typically relied upon in behaviour analytic studies due to their poor predictive validity and susceptibility to influence from extraneous social and verbal processes. In contrast, as

an index of enjoyment, Electro-dermal Activity (EDA) are unhampered by social or verbal processes. Thus, as EDA is a more direct measure of the correlates of enjoyable game play, it should function as a more reliable measure of future choice than a verbal report. In effect, given the findings of Chapter 4 researchers may find it more beneficial to employ a functional-analytic psychophysiological approach to game choice than to use verbal reports for such purposes.

Importantly, the particular technique employed for analysing EDA data in Chapter 4 appears to circumvent two problems which had previously lead researchers to regard EDA as unsuitable for the analysis of game play (see Sykes & Brown, 2003). Sykes and Brown pointed out that increased perspiration and muscle tension necessarily accompany fast paced arcade-style games and that thermoregulatory activity and movement artifacts interfere with reliable EDA measurement. Thus, it would appear that EDA could not be used to reliably measure participants' arousal levels during game play. However, the technique used to analyse Skin Conductance Level (SCL) data in Chapter 4 circumvents Sykes and Browns' concerns regarding movement artifacts by examining tonic rather than phasic SCL. As such, the individual peaks and troughs in SCL associated with movement artifacts do not contribute significantly to the data observed. In addition, concerns over the fact that increased perspiration would lead to unreliable SCL measurement are unfounded, due to electrode placement. Thermoregulatory sweating on the palmer surfaces of the human body by both eccrine and apocrine glands rarely occurs (Fowles, 1986; Jakubovic & Ackerman, 1985). So, despite the concerns raised by Sykes and Brown, the particular technique developed in the current chapter for analysing SCL data appears to have yielded potentially important findings, suggesting that electrodermal activity is a sensitive and useful measure of behaviour during game play.

Interestingly, it appears, from studies reported in Chapter 4, that computer games could be conceived of as a type of sustained attention task. A large body of research exists that identifies a gradual increase in physiological arousal as an accompaniment of sustained attention (see Bohlin, 1976; Grim, 1967; O'Connell, Bellgrove, Dockree, & Robertson, 2004; Oken, Salinsky, & Elsas, 2006; Smallwood, Davies, Heim, Finnigan, Sudberry, O'Connor, & Obonsawin, 2004). Precisely this pattern of arousal was observed in the current studies during games that participants both chose to play later and rated as enjoyable. Thus, physiological data suggests that the games that participants enjoyed were the games which sustained participants' attention. These findings provide a possible conceptual bridge between the current behavioural research on enjoyment in computer games, and a standard and well-researched concept in cognitive psychology. It was suggested in Chapter 4 that both fields may benefit from cross-fertilisation of these obviously parallel ideas; that sustained attention and the potency of reinforcing experiences (i.e., enjoyment) may be closely related.

Experiments 4.1 and 4.3 provided specific slope-of-graph gradients that may be useful as an absolute coefficient for indicating future game choices, if not enjoyment. In other words, Experiments 4.1 and 4.3 found that it was the slope of the arousal increase across time, rather than absolute levels of arousal that provided the best indicators of enjoyment of a game. Future research with a wide variety of different games may help to further determine whether the SCL slopes observed here for enjoyable and chosen games are applicable to a wider variety of games. Future research may also investigate the pattern of arousal that would be observed if subjects were exposed to a game over longer periods of time.

The findings of Chapter 4 are of immediate practical importance to computer games researchers and designers. These experiments demonstrate that SCL can be used to measure and predict participants' preference for particular games. More specifically, a signature SCL trend was identified that appears to describe game enjoyment and predict game choice. Additionally, a distinct SCL trend was identified that describes dislike for a game and predicts that the game will not be chosen later to be played. Using this information, it now seems possible to build a game that is adaptive to participants' game preference and ability levels. SCL technology is cheap, can be attached to the body non-intrusively, can be sold with a game, and interfaced with the game easily, as demonstrated by the success of *The Journey to Wild Devine* (© Wild Devine, 2006). Thus, games could be created that are sold with an optional SCL add-on that tracks SCL during game play, and searches for the two SCL signatures identified in the current research. This information could then serve as an input to the games, so that complexity levels are maintained so long as the signature enjoyment pattern of gradually increasing SCL is observed. Complexity could then be increased if the SCL trend stabilises, or shows a decrease across time during game-play. If the signature can not be re-established by increasing complexity, complexity could be decreased until the signature pattern is re-established. Thus, it appears that the current research has laid the technical and conceptual groundwork for a whole new set of research opportunities within the gaming and human-computer interaction fields that may well lead to a new and exciting generation of adaptive computer games, currently just beyond the horizon of game developers.

5.5 Concluding Comments

The current research programme represents a first step in the psychological analysis of on-line game playing, a multi-billion dollar worldwide industry. Network Latency and 'game challenge' were identified as two important variables in participants' enjoyment of on-line games and the current

research agenda adopted a behaviour analytic methodology to studying the effects of these variables. The experiments presented in Chapter 2 defined 'game challenge' in terms of derived relational responding (or Relational Complexity) and found that participants were able to consistently respond in accordance with stimulus generalization and both one and three-node stimulus equivalence relations in the context of a computer game. The presence of Network Latency in a game was found detrimental to the game playing experience, but increasing the length of those delays had no significant effect on either verbal reports or behavioural measures. Relational Complexity had no significant effect on results in three of the four experiments presented in Chapter 2, suggesting that the games presented were not sufficiently complex to be interesting to participants. The experiments presented in Chapter 3 defined 'game challenge' in terms of more complex forms of derived relational responding and found that participants were able to consistently respond in accordance with derived 'SAME' and 'OPPOSITE' relations in the context of a computer game. As in Chapter 2, the presence of Network Latency in a game was found detrimental to the game playing experience, but once again increasing the length of those delays had no significant effect on either verbal reports or behavioural measures. Participants were more successful at the relationally less complex levels of the games examined in Chapter 3 and demonstrated preference for those levels. The experiments reported in Chapter 4 were conducted in order to develop novel behavioural and physiological measures of enjoyment in game playing. Game challenge was defined in terms of speed of stimulus presentation and the number of stimuli presented. It was found that participants' preference for games of varying difficulty was dependent on their experience with those games. Across repeated presentations, preference shifted from simpler to more complex games. In addition, a method was developed for analysing EDA, which successfully differentiated games on the basis of the preference shown for them by participants. This technology may be of immediate practical application within the computer games industry.

The current research would appear to make a contribution to the gaming industry in providing some exciting findings and lines of enquiry for further research into the effects of Network Latency and game challenge on gaming behaviour and enjoyment. In addition, a novel technology for the analysis of EDA data as a measure of enjoyment in gaming was developed that may have practical applications within the computer games industry. The current research also contributes to the literature on derived stimulus relations, in showing that this concept may be extended to the understanding of online game playing; a highly novel aspect of human behaviour. Finally, while many questions remain regarding all of the issues explored in the current research, this is to be expected. No experimental programme can realistically expect to conduct the *experimentum crucis*

(the experiment that will finally answer a crucial question relevant to the field). Rather, a significant contribution of any research programme is that it might lay the groundwork for further research that may help to successively approximate more or less useful conclusions regarding a research question. This process of building knowledge systems from the ground up is as important as the acquisition of research findings themselves. Insofar as the current research has provided both empirical insights and research method developments, therefore, it would appear to represent a contribution to the research fields of derived relational responding, behavior analysis, and game development research more generally.

References

- Aarseth, E. (2001). Computer game studies, year one. *The International Journal of Computer Game Research*, 1(1), <http://www.gamestudies.org/0101/editorial.html>.
- Adams, B. J., Fields, L., & Verhave (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record*, 43, 133–152.
- Adams, E. (2001a). Part one: Narrative, gamasutra.com. Retrieved March 4, 2006, from http://www.gamasutra.com/features/20010521/adams_01.htm.
- Adams, E. (2001b). Replayability part 2: Game mechanics, gamasutra.com. Retrieved March 4, 2006, from http://www.gamasutra.com/features/20010703/adams_02.htm.
- Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers*, 16(5), 857–878.
- Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, 47(2), 309–320.
- Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record*, 50, 603–628.
- Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record*, 44, 91–124.
- Barnes, D., Browne, M., Smeets, P., & Roche, B. (1995). A transfer of functions and a conditional transfer of functions through equivalence relations in three to six year old children. *The Psychological Record*, 45, 405–430.
- Barnes, D., & Hampson, P. (1993). Stimulus equivalence and connectionism: Implications for behavior analysis and cognitive science. *The Psychological Record*, 43, 617–638.
- Barnes, D., & Keenan, M. (1989). Instructed human fixed-interval performance: The effects of the experimental setting. *The Psychological Record*, 39, 351–364.

REFERENCES

- Barnes, D., & Keenan, M. (1993). Concurrent activities and instructed human fixed-interval performance. *Journal of the Experimental Analysis of Behavior*, *59*, 501–520.
- Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes. *The Psychological Record*, *50*, 251–265.
- Barnes-Holmes, D., Barnes-Holmes, Y., Smeets, P. M., Cullinan, V., & Leader, G. (2004). Relational frame theory: Conceptual and procedural issues. *International Journal of Psychology and Psychological Therapy*, *4*, 181–214.
- Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and skinner's verbal behavior: A possible synthesis. *The Behavior Analyst*, *23*, 69–84.
- Barnes-Holmes, D., D. O'Hora, D., Roche, B., Hayes, S., Bisset, R., & Lyddy, F. (2001a). Understanding and verbal regulation. In S. Hayes, D. Barnes-Holmes, & B. T. Roche (Eds.) *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*, (pp. 103–118). New York: Plenum.
- Barnes-Holmes, D., Hayes, S., & Dymond, S. (2001b). Self and self-directed rules. In S. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational frame theory: A post-skinnerian account of human language and cognition*, (pp. 119–140). New York: Plenum Press.
- Barnes-Holmes, Y., Barnes-Holmes, D., & Cullinan, V. (2001c). Education. In S. Hayes, D. Barnes-Holmes, & B. T. Roche (Eds.) *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*, (pp. 181–197). New York: Plenum.
- Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. (2001d). Exemplar training and a derived transformation of function in accordance with symmetry. *The Psychological Record*, *51*, 287–308.
- Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001e). Exemplar training and a derived transformation of function in accordance with symmetry ii. *The Psychological Record*, *51*, 589–603.
- Baron, A. (1999). Statistical inference in behavior analysis: friend or foe? *The Behavior Analyst*, *22*(2), 83–85.
- Bauchhage, C., Thureau, C., & Sagerer, G. (2003). Learning human-like opponent behavior for interactive computer games. In B. Michaelis, & G. Krell (Eds.) *Pattern Recognition*, Lecture Notes in Computer Science, (pp. 148–155). Springer Berlin / Heidelberg.

REFERENCES

- Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for a relational operant. *Journal of Applied Behavior Analysis, 40*(1), 45–71.
- Bersak, D., McDarby, G., Augenblick, N., McDarby, P., McDonnell, D., McDonald, B., & Karkun, R. (2001). Intelligent biofeedback using an immersive competitive environment. In *Online Proceedings for the Designing Ubiquitous Computing Games Workshop, UbiComp 2001, Atlanta GA*.
- Boe, R., & Winokur, S. (1978). A procedure for studying echoic control in verbal behavior. *Journal of the Experimental Analysis of Behavior, 30*(2), 213–217.
- Bohlin, G. (1976). Delayed habituation of the electrodermal orienting response as a function of increased level of arousal. *Psychophysiology, 13*, 345–351.
- Branch, M. (1999). Statistical inference in behavior analysis: some things significance testing does and does not do. *The Behavior Analyst, 22*(2), 87–92.
- Buckley, S. (2007). Latency-free gaming. Retrieved January 16, 2008, from http://www.telecommagazine.com/search/article.asp?HH_ID=AR_3400.
- Burke, T. (2005). Can a table stand on one leg? critical and ludological thoughts on star wars: Galaxies. *The International Journal of Computer Game Research, 1*, Retrieved January 16, 2008, from <http://www.gamestudies.org/0501/burke/>.
- Cabello, F., Barnes-Holmes, D., O'Hora, D., & Stewart, I. (2002). Using visual basic in the experimental analysis of human behavior: A brief introduction. *Experimental Analysis of Human Behavior Bulletin, 20*, 17–20.
- Cabello, F., Luciano, C., Gomez, I., & Barnes-Holmes, D. (2004). Human schedule performance, protocol analysis, and the "silent dog" methodology. *The Psychological Record, 54*(3), 405–422.
- Cabello, F., & O'Hora, D. (2002). Addressing the limitations of protocol analysis in the study of complex human behavior. *International Journal of Psychology and Psychological Therapy, 2*(2), 115–130.
- Cahill, B., Ward, R., Marsden, P., & Johnson, C. (2001). Fun, work and affective computing: can psychophysiology measure fun? *Interfaces 46, Spring*, 11.
- Case, D. A., Ploog, B. O., & Fantino, E. (1990). Observing behavior in a computer game. *Journal of the Experimental Analysis of Behavior, 54*(3), 185–199.

REFERENCES

- Castronova, E. (2003). On virtual economies. *The International Journal of Computer Game Research*, 3(2), Retrieved January 16, 2008, from <http://www.gamestudies.org/0302/castronova/>.
- Catania, C. A. (1998). *Learning*, 4 ed. Cornwall-on-Hudson, NY: Sloan Publishing.
- Chase, P. N., & Danforth, J. S. (1991). The role of rules in concept learning. In L. J. Hayes, & P. N. Chase (Eds.) *Dialogues on verbal behavior*, (pp. 205–222). Reno, NV: Context Press.
- Collet, C., Vernet-Maury, E., Delhomme, G., & Dittmar, A. (2005). Autonomic nervous system response patterns specificity to basic emotions. *Journal of the Autonomic Nervous System*, 62(1–2), 45–57.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2006). *Applied Behavior Analysis*, 2 ed. NJ: Prentice Hall.
- Coyne, R. (2003). Mindless repetition: Learning from computer games. *Design Studies*, 24(3), 199–212.
- Critchfield, T., & Epting, L. (1998). The trouble with babies and the value of bathwater: complexities in the use of verbal reports as data. *The Analysis of Verbal Behavior*, 15, 65–74.
- Davis, J. P., Steury, K., & Pagulayan, R. J. (2005). A survey method for assessing perceptions of a game: The consumer playtest in game design. *The International Journal of Computer Game Research*, 5(1), Retrieved January 16, 2008, from http://www.gamestudies.org/0501/davis_steury_pagulayan/.
- Dawson, M., Schell, A., & Filion, D. (2000). The electrodermal system. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.) *Handbook of Psychophysiology*, (pp. 200–223). Cambridge: Cambridge University Press, 2 ed.
- De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *The Quarterly Journal of Experimental Psychology*, 55, 965–985.
- DeHouwer, J., Crombez, G., & Baeyens, F. (2005). Avoidance behavior can function as a negative occasion setter. *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 101–106.
- Delaney, D., Meenaghan, P., Ward, T., & McLoone, S. (2004). Examining user performance in the presence of latency and jitter in distributed interactive applications. In *ISSC 2004*. Belfast, Ireland.

REFERENCES

- Delaney, D., Ward, T., & McLoone, S. (2003). Reducing update packets in distributed interactive applications using a hybrid model. In *Proceedings of the 16th ISCA International Conference on Parallel and Distributed Computing Systems (PDCS)*, (pp. 417–422). Reno, NV.
- Desurvire, H., Caplan, M., & Toth, J. A. (2004). Using heuristics to evaluate the playability of games. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, (pp. 1509–1512). New York, NY, USA: ACM.
- Dixon, K., Malak, R., & Khosla, P. (2000). Incorporating prior knowledge and previously learned information into reinforcement learning agents. Tech. rep., Carnegie Mellon University, Institute for Complex Engineered Systems.
- Dixon, M., Marley, J., & Jacobs, E. (2003a). Delay discounting by pathological gamblers. *Journal of Applied Behavior Analysis*, *36*, 449–458.
- Dixon, M. R., Dymond, S., Rehfeldt, R. A., Roche, B., & Zlomke, K. R. (2003b). Terrorism and relational frame theory. *Behavior and Social Issues*, *12*, 129–147.
- Dixon, M. R., & MacLin, O. H. (2003). *Visual basic for behavioral psychologists*. Reno, NV:: Context Press.
- Dougher, M., Auguston, E., Markham, M., Greenway, D., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, *62*, 331–351.
- Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behavior*, *64*(2), 163–184.
- Dymond, S., & Barnes, D. (1997). Behavior-analytic approaches to self-awareness. *The Psychological Record*, *47*(2), 181–200.
- Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2007). A transformation of avoidance responses in accordance with same and opposite relational frames. *Journal of the Experimental Analysis of Behavior*, *88*, 249–262.
- Dymond, S., Roche, B., & Rhefeldt, R. (2005). Relational frame theory and the transformation of stimulus functions. *Revista Latinamericana De Psicologia*, *37*(2), 291–303.
- Edelberg, R. (1972). Electrical activity of the skin: Its measurement and uses in psychophysiology. In N. G. . R. Sternbach (Ed.) *Handbook of Psychophysiology*, (pp.

REFERENCES

- 367–418). New York: Holt.
- Ericsson, A. K., & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
- Fabricatore, C., Nussbaum, M., & Rosas, R. (2002). Playability in action videogames: A qualitative design model. *Human-Computer Interaction, 17*(4), 311–368.
- Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology, 56*(2), 171–184.
- Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53*(3), 345–358.
- Fields, L., Reeve, K., Rosen, D., Varelas, A., & Adams, B. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior, 67*(3), 367–389.
- Fowles, D. G. (1986). The eccrine system and electrodermal activity. In M. G. H. Coles, B. Donchin, & S. W. Porges (Eds.) *Psychophysiology: Systems, processes, and applications*, (pp. 51–96). New York: Guilford Press.
- Fung, M. T., Raine, A., Loeber, R., Lynam, D. R., Steinhauer, S. R., Venables, P. H., & Stouthamer-Loeber, M. (2005). Reduced electrodermal activity in psychopathy-prone adolescents. *Journal of Abnormal Psychology, 114*(2), 187–196.
- Galizio, M. (2004). Relational frames: Where do they come from? a comment on barnes-holmes and hayes (2003). *The Behavior Analyst, 27*, 107–112.
- Gilleade, K., & Dix, A. (2004). Using frustration in the design of adaptive videogames. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology, singapore, ACM International Conference Proceeding Series*, vol. 74, (pp. 228–232).
- Gingold, C. (2005). What warioware can teach us about game design. *The International Journal of Computer Game Research, 5*(1), Retrieved January 16, 2008, from <http://www.gamestudies.org/0501/gingold/>.
- Gómez, S., López, F., Martín, C. B., Barnes-Holmes, Y., & Barnes-Holmes, D. (2007).

REFERENCES

- Exemplar training and a derived transformation of functions in accordance with symmetry and equivalence. *The Psychological Record*, 57(2), 273–293.
- Grim, F., Paul (1967). A sustained attention comparison of children and adults using reaction time set and the gsr. *Journal of Experimental Child Psychology*, 5(1), 26–38.
- Hashimoto, Y., & Ishibashi, Y. (2006). Influences of network latency on interactivity in networked rock-paper-scissors. In *NetGames '06: Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games*, (p. 23). New York, NY, USA: ACM.
- Hayes, S. (1986). The case of the silent dog - verbal reports and the analysis of rules. a review of ericsson and simon's protocol analysis: Verbal reports as data. *Journal of the Experimental Analysis of Behavior*, 45, 351–363.
- Hayes, S. C. (Ed.) (1989). *Rule-Governed Behavior: Cognition, Contingencies, & Instructional Control*. Context Press.
- Hayes, S. C. (1993). Why environmentally based analyses are necessary in behavior analysis. *Journal of the Experimental Analysis of Behavior*, 60(2), 461–463.
- Hayes, S. C. (1994). Relational frame theory: A functional approach to verbal events. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.) *Behavior Analysis of Language and Cognition*, (pp. 9–30). Reno, NV: Context Press.
- Hayes, S. C., & Barnes-Holmes, D. (2004). Relational operants: processes and implications: a response to palmer's review of relational frame theory. *Journal of the Experimental Analysis of Behavior*, 82(2), 213–24; discussion 225–34.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.) (2001a). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.
- Hayes, S. C., Blackledge, J. T., & Barnes-Holmes, D. (2001b). Language and cognition: Constructing an alternative approach within the behavioral tradition. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational frame theory: A post-Skinnerian account of human language and cognition*, (pp. 3–20). New York: Plenum.
- Hayes, S. C., Gifford, E., Townsend, R. J., & Barnes-Holmes, D. (2001c). Thinking, problem solving, and pragmatic verbal analysis. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational frame theory: A post-Skinnerian account of human language and cognition*, (pp. 87–103). New York: Plenum.
- Hayes, S. C., & Wilson, K. G. (1996). Criticisms of relational frame theory: Implications for a

REFERENCES

- behavior analytic account of derived stimulus relations. *The Psychological Record*, *46*, 221–236.
- Healy, O., Barnes, D., & Smeets, P. (1998). Derived relational responding as an operant: The effects of between-session feedback. *The Psychological Record*, *48*, 511–536.
- Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*(2), 207–227.
- Herbelin, B., Benzaki, P., Riquier, F., Renault, O., & Thalmann, D. (2004). Using physiological measures for emotional assessment: a computer-aided tool for cognitive and behavioural therapy. In *Proc. 5th Intl Conf. Disability, Virtual Reality & Assoc. Tech.*. Oxford, UK.
- Holland, J. (1958). Counting by humans on fixed-ratio schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *1*, 179–181.
- Ishibashi, Y., Nagasaka, M., & Fujiyoshi, N. (2006). Subjective assessment of fairness among users in multipoint communications. In *ACE '06: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, (p. 69). New York, NY, USA: ACM.
- Jakubovic, H. R., & Ackerman, A. B. (1985). Structure and function of the skin, section 1: Development, morphology, and physiology. In S. L. Moschella, & H. J. Hurley (Eds.) *Dermatology*, *1*, (pp. 1–74). Philadelphia, PA: Saunders.
- Jennings, A. (2002). Creating an interactive science murder mystery game: the optimal experience of flow. *Professional Communication, IEEE Transactions on*, *45*(4), 297–301.
- Jørgensen, K. (2003). Problem solving: The essence of player action in computer games. In *Digital Games Research Conference*. The Netherlands: University of Utrecht.
- Juul, J. (2003a). The game, the player, the world: Looking for a heart of gameness. In M. Copier, & J. Raessens (Eds.) *Level Up: Digital Games Research Conference Proceedings*, (pp. 30–45). Utrecht University.
- Juul, J. (2003b). Just what is it that makes computer games so different, so appealing? Column for the IGDA (international game developers association) Retrieved January 16, 2008, from http://www.igda.org/columns/ivorytower/ivory_Apr03.php.
- Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method.

REFERENCES

- Science*, 306 (5702), 1776–1780.
- Kiili, K. (2005a). Content creation challenges and flow experience in educational games: The it-emperor case. *The Internet and Higher Education*, 8(3), 183–198.
- Kiili, K. (2005b). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, 8(1), 13–24.
- Klein, J., Moon, Y., & Picard, R. (2002). This computer responds to user frustration: Theory, design, and result. *Interacting with Computers*, 14(2), 119–140.
- Krauss, M., Scheuchenpflug, R., Piechulla, W., & Zimmer, A. (2001). Measurement of presence in virtual environments. In K. L. A.C. Zimmer (Ed.) *Experimentelle Psychologie im Spannungsfeld von Grundlagenforschung und Anwendung. Proceedings 43. Tagung experimentell arbeitender Psychologen (CD-ROM)*. Regensburg: Universitätsbibliothek. ISBN: 3-88246-225-6.
- Lacey, J., Kagan, J., Lacey, B., & Moss, H. (1963). The visceral level: situational determinants and behavioral correlates of autonomic response patterns. In P. Knapp (Ed.) *Expression of Emotions in Man*, vol. 1963, (pp. 161–196). New York: International University Press.
- Lamarre, J., & Holland, J. (1985). The functional independence of mands and tacts. *Journal of the Experimental Analysis of Behavior*, 43, 5–19.
- LaPierre, R. T. (1934). Attitude versus actions. *Social Forces*, 13, 230–237.
- Leander, J., Lippman, L., & Meyer, M. (1968). Fixed interval performance as related to instructions and subjects verbalizations of the reinforcement contingency. *The Psychological Record*, 18, 469–474.
- Lee, V. (1981). Prepositional phrases spoken and heard. *Journal of the Experimental Analysis of Behavior*, 35, 227–242.
- Lee, V., & Pegler, A. (1982). Effects on spelling of training children to read. *Journal of the Experimental Analysis of Behavior*, 37, 311–322.
- Leung, J.-P. (1989). Psychological distance to reward: A human replication. *Journal of the Experimental Analysis of Behavior*, 51(3), 343–352.
- Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, 45(3), 587–597.
- Levenson, R. W., & Gottman, J. M. (1985). Physiological and affective predictors of change in relationship satisfaction. *Journal of Personality and Social Psychology*, 49(1), 85–94.

REFERENCES

- Levinson, D., Edelberg, R., & Bridger, W. (1984). The orienting response in schizophrenia: Proposed resolution of a controversy. *Biological Psychiatry, 19*, 489–507.
- Lipkens, G., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of derived stimulus relations in an infant. *Journal of Experimental Child Psychology, 56*, 201–239.
- Loftus, G. R., & Loftus, E. F. (1983). *Mind at Play: The Psychology of Video Games*. New York, NY: Basic Books.
- Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning and Memory, 11*(5), 495–500.
- Lowe, C., Harzem, P., & Hughes, S. (1978). Determinants of operant behavior in humans: some differences from animals. *Quarterly Journal of Experimental Psychology, 30*, 373–386.
- MacLin, O. H., Dixon, M. R., & Hayes, L. J. (1999). A computerized slot machine simulation to investigate the variables involved in gambling behavior. *Behavior Research Methods, Instruments, & Computers, 31*(4), 731–734.
- Malone, T., & Lepper, M. (1987). Making learning fun: a taxonomy of intrinsic motivation for learning. In R. Snow, & M. Farr (Eds.) *Aptitude, Learning, and Instruction*, vol. 3 of *Cognitive and Affective Process Analysis*, (pp. 223–253). Hillsdale, NJ: Erlbaum.
- Malone, T. W. (1982). Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 conference on Human factors in computing systems*, (pp. 63–68). New York, NY, USA: ACM.
- Mandryk, R. L., & Inkpen, K. M. (2004). Physiological indicators for the evaluation of co-located collaborative play. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, (pp. 102–111). New York, NY, USA: ACM.
- Manninen, T., & Kujanpaa, T. (2005). The hunt for collaborative war gaming - case: Battlefield 1942. *The International Journal of Computer Game Research, 5*(1), Retrieved January 16, 2008, from http://gamestudies.org/0501/manninen_kujanpaa/.
- McIlvane, W. J. (2003). A stimulus in need of a response (review of relational frame theory: A post-skinnerian account of human language and cognition). *Analysis of Verbal Behavior, 19*, 29–37.
- Morlock, H., Yando, T., & Nigolean, K. (1985). Motivation of video game players. *Psychological Reports, 57*, 247–250.

REFERENCES

- Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2005). Derived manding in children with autism: synthesizing skinner's verbal behavior with relational frame theory. *Journal of Applied Behavior Analysis, 38*(4), 445–462.
- Nasoz, F., Alvarez, K., Lisetti, L., & Finkelstein, N. (2003). Emotion recognition from physiological signals using wireless sensors for presence technologies. *International Journal of Cognition, Technology, and Work, 6*(1), 4–14.
- O'Connell, R. G., Bellgrove, M. A., Dockree, P. M., & Robertson, I. H. (2004). Reduced electrodermal response to errors predicts poor sustained attention performance in attention deficit hyperactivity disorder. *Neuroreport, 15*(16), 2535–2538.
- O'Hora, D., Barnes-Holmes, D., Roche, B., & Smeets, P. (2004). Derived relational networks as novel instructions: A possible model of generative verbal control. *The Psychological Record, 54*, 437–460.
- O'Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record, 52*, 51–75.
- Oken, B., Salinsky, M., & Elsas, S. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology, 117*, 1885–1901.
- Palmer, D. (2004). Generic response classes and relational frame theory: response to hayes and barnes-holmes. *Journal of the Experimental Analysis of Behavior, 82*(2), 225–234.
- Pantel, L., & Wolf, L. C. (2002). On the impact of delay on real-time multiplayer games. In *NOSSDAV '02: Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video*, (pp. 23–29). New York, NY, USA: ACM.
- Parente, A., & Parente, R. (2006). Mind-operated devices: Mental control of a computer using biofeedback. *CyberPsychology & Behavior, 9*(1), 1–4.
- Perone, M. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: Use and abuse of subjects' verbal reports. *The Behavior Analyst, 11*, 71–75.
- Perone, M. (1999). Statistical inference in behavior analysis: experimental control is better. *The Behavior Analyst, 22*(2), 109–116.
- Picard, R. W. (1997). *Affective computing*. Cambridge, MA, USA: MIT Press.

REFERENCES

- Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *Int. J. Hum.-Comput. Stud.*, *62*(2), 231–245.
- Raia, C. P., Shillingford, S. W., Miller, H. L., & Baier, P. S. (2000). Interaction of procedural factors in human performance on yoked schedules. *Journal of the Experimental Analysis of Behavior*, *74*(3), 265–281.
- Roche, B., & Barnes, D. (1995a). The establishment and electrodermal assessment of conditioned sexual responses. *Experimental Analysis of Human Behavior Bulletin*, *13*, 26–29.
- Roche, B., & Barnes, D. (1995b). Measuring ill-defined events...and other problems. a reply to augustson. *Experimental Analysis of Human Behavior Bulletin*, *13*, 31–32.
- Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior*, *67*(3), 275–301.
- Roche, B., Linehan, C., Ward, T., Dymond, S., & Reheldt, R. (2004). The unfolding of the relational operant: A real-time analysis using electroencephalography and reaction time measures. *International Journal of Psychology and Psychological Therapy*, *4*, 1–17.
- Rollings, A., & Adams, E. (2003). *Andrew Rollings and Ernest Adams on Game Design*. Boston: New Riders Publishing.
- Sakurazawa, S., Yoshida, N., Munekata, N., Omi, A., Takeshima, H., Koto, H., Gentsu, K., Kimura, K., Kawamura, K., Miyamoto, M., Arima, R., Mori, T., Sekiya, T., Furukawa, T., Hashimoto, Y., & Numata, H. (2003). A computer game using galvanic skin response. In *ACM International Conference Proceeding Series; Vol. 38, Proceedings of the second international conference on Entertainment computing, Pittsburgh, Pennsylvania*, (pp. 1–3).
- Salen, K., & Zimmerman, E. (2003). *Rules of play game design fundamentals*. Cambridge, Mass.: MIT Press.
- Salzinger, K. (1958). A method of analysis of the process of verbal communication between a group of emotionally disturbed adolescents and their friends and relatives. *Journal of Social Psychology*, *47*, 39–53.
- Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record*, *43*, 725–744.

REFERENCES

- Scheirer, J., Fernandez, R., Klein, J., & Picard, R. (2002). Frustrating the user on purpose: A step toward building an affective computer. *Interact Comput*, *14*(2), 93–118.
- Schlenker, R., Cohen, R., Hubmann, W., Mohr, F., Wahlheim, C., Watzl, H., & Werther, P. (1995). Electrodermal and vascular orienting response in schizophrenic patients: Relationship to symptoms and medication. *European Archives of Psychiatry and Clinical Neuroscience*, *245*(3), 152–158.
- Sharkey, P. M., Ryan, M. D., & Roberts, D. J. (1998). A local perception filter for distributed virtual environments. In *VRAIS '98: Proceedings of the Virtual Reality Annual International Symposium*, (p. 242). Washington, DC, USA: IEEE Computer Society.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, *14*, 5–13.
- Sidman, M. (1994). *Equivalence Relations and Behavior: A Research Story*. Boston, MA: Authors Cooperative, Inc.
- Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior*, *37*(1), 23–44.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, *37*(1), 5–22.
- Skinner, B. F. (1953). *Science and Human Behavior*. New York: The Free Press.
- Skinner, B. F. (1957). *Verbal behaviour*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1959). *Cumulative Record*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1974). *About Behaviorism*. New York: Random House.
- Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*(4), 657–690.
- Smeets, P. M., & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes*, *69*, 281–293.

REFERENCES

- Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, *56*(3), 519–555.
- Stevens, M., Moget, P., De Greef, M., Lemmink, K., & Rispen, P. (2000). The groningen enjoyment questionnaire: A measure of enjoyment in leisure-time physical activity. *Perceptual and Motor Skills*, *90*(2), 601–604.
- Stewart, I., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2002). A functional-analytic model of analogy: A relational frame analysis. *Journal of the Experimental Analysis of Behavior*, *78*, 375–396.
- Stewart, I., Barnes-Holmes, D., S.C, H., & Lipkens, R. (2001). Relations among relations: Analogies, metaphors, and stories. In S. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. New York, NY: Kluwer Academic/Plenum Publishers.
- Suits, B. (1978). *The Grasshopper: Games, Life and Utopia*. Toronto, Canada: University of Toronto Press.
- Sykes, J., & Brown, S. (2003). Affective gaming: measuring emotion through the gamepad. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, (pp. 732–733). New York, NY, USA: ACM.
- Tonneau, F. (2001a). Equivalence relations: A critical analysis. *European Journal of Behavior Analysis*, *2*, 1–33.
- Tonneau, F. (2001b). Equivalence relations: A reply. *European Journal of Behavior Analysis*, *2*, 99–128.
- Turpin, G., & Clements, K. (1993). Electrodermal activity and psychopathology: The development of the palmar sweat index as an applied measure for use in clinical settings. In J. Roy, W. Boucsein, D. Fowles, & J. Gruzelier (Eds.) *Electrodermal Activity: From Physiology to Psychology*, (pp. 49–60). New York: Plenum Press.
- Vaghi, I., Greenhalgh, C., & Benford, S. (1999). Coping with inconsistency due to network delays in collaborative virtual environments. In *VRST '99: Proceedings of the ACM symposium on Virtual reality software and technology*, (pp. 42–49). New York, NY, USA: ACM.
- Verwey, W., & Veltman, H. (1996). Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, *2*(3), 270–285.

REFERENCES

- Voiskounsky, A., Mitina, O., & Avetisova, A. (2004). Playing online games: Flow experience. *PsychNology journal*, 2(3), 259–281.
- Vorderer, P., Hartmann, T., & Klimmt, C. (2003). Explaining the enjoyment of playing video games: the role of competition. In *ICEC '03: Proceedings of the second international conference on Entertainment computing*, (pp. 1–9). Pittsburgh, PA, USA: Carnegie Mellon University.
- Vyzas, E. (1999). *Recognition of Emotional and Cognitive States Using Physiological Data*. Ph.D thesis, MIT, Cambridge, MA.
- Whelan, R., & Barnes-Holmes, D. (2004). A derived transformation of consequential functions in accordance with the relational frames of same and opposite. *Journal of the Experimental Analysis of Behavior*, 82, 177–195.
- Whelan, R., Barnes-Holmes, D., & Dymond, S. (2006). The transformation of consequential functions in accordance with the relational frames of more-than and less-than. *Journal of the Experimental Analysis of Behavior*, 86, 317–335.
- Williams, R., & Clippinger, C. (2002). Aggression, competition and computer games: computer and human opponents. *Computers in Human Behavior*, 18(5), 495–506.
- Wilson, K., Hayes, S., Gregg, J., & Zettle, R. (2001). Psychopathology and psychotherapy. In S. Hayes, D. Barnes-Holmes, & B. Roche (Eds.) *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*, (pp. 211–239). New York, NY: Kluwer Academic/Plenum Publishers.
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240.
- Woods, S. (2004). Loading the dice: The challenge of serious videogames. *The International Journal of Computer Game Research*, 4(1), Retrieved January 16, 2008, from <http://www.gamestudies.org/0401/woods/>.
- Zahn, T. P., Grafman, J., & Tranel, D. (1999). Frontal lobe lesions and electrodermal activity: effects of significance. *Neuropsychologia*, 37, 1227–1241.

Appendix A

The Day Reconstruction Method (DRM)

A.1 Instrument Documentation

- Daniel Kahneman, Princeton University. *kahneman@princeton.edu*
- Alan B. Krueger, Princeton University. *akrueger@princeton.edu*
- David Schkade, University of Texas, Austin. *schkade@mail.utexas.edu*
- Norbert Schwarz, University of Michigan. *nschwarz@umich.edu*
- Arthur Stone, Stony Brook University. *astone@mail.psychiatry.sunysb.edu*

How people spend their time and how they experience the various activities and settings of their lives are significant questions for researchers in diverse disciplines. The Day Reconstruction Method (DRM) is designed to collect data describing the experiences a person has on a given day, through a systematic reconstruction conducted on the following day. The DRM builds on the strengths of time-budget measurement (Juster & Stafford, 1985; Robinson & Godbye, 1997) and experience sampling (Stone, Shiffman, & DeVries, 1999), and employs techniques grounded in cognitive science. The conceptual rationale and illustrative findings, based on a sample of employed women in Texas, are presented in Kahneman, Krueger, Schkade, Schwarz, and Stone (2004). The results indicate a close correspondence between the DRM and established results from experience sampling.

Key advantages of the DRM include:

- Joint assessment of activities and subjective experiences
- Information about the duration of each experience, allowing for duration weighted analyses of experiences
- Lower respondent burden than typical for experience sampling methods
- More complete coverage of the day than typical for experience sampling methods
- Lower susceptibility to retrospective reporting biases than typical for global reports of daily experiences
- High flexibility in adapting the content of the instrument to the needs of the specific study

A.1.1 Instruments

The DRM asks respondents to reconstruct the previous day by completing a structured self-administered questionnaire. A respondent first reinstantiates the previous day into working memory by producing a short diary consisting of a sequence of episodes. This instrument is documented as “Packet 2.” Its format draws on insights from cognitive research with Event History Calendars (Belli, 1998) and facilitates retrieval from autobiographical memory through multiple pathways. Its episodic reinstantiation format attenuated biases commonly observed in retrospective reports (Robinson & Clore, 2002; Schwarz & Oyserman, 2001; Schwarz & Sudman, 1994). Respondents’ diary entries are confidential and the diary does not need to be returned to the researcher. This allows respondents to use idiosyncratic notes, including details they may not want to share. Next, respondents receive a response form (documented as “Packet 3”) and are encouraged to draw on their confidential diary notes to answer a series of questions. These questions ask them to describe key features of each episode, including (1) when the episode began and ended, (2) what they were doing, (3) where they were, (4) whom they were interacting with, and (5) how they felt on multiple affect dimensions. This response form is returned to the researcher for analysis. In addition, respondents answer a number of questions about themselves and the circumstances of their lives (e.g., demographics, job characteristics, personality measures). “Packet 1” and “Packet 4” document the variables assessed in the study reported in Kahneman et al. (2004). DRM Documentation - 4

A.1.2 Administration

For methodological reasons, it is important that respondents complete the diary before they are aware of the specific content of the later questions about each episode. Early knowledge of these

questions may affect the reconstruction of the previous day and may introduce selection biases. This is best achieved by presenting the diary (“Packet 2”) and the response form (“Packet 3”) in separate envelopes, asking respondents not to open the next envelope until the previous material is completed. The DRM can be administered individually or in group settings. In our experience, adults from the general population can complete the full set of materials in 45 to 75 minutes.

A.1.3 References

- Belli, R. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys, *Memory*, 6, 383-406.
- Juster, F. T., & Stafford, F. P. (eds.) (1985). *Time, goods, and well-being*. Ann Arbor, MI: Institute for Social Research Press.
- Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The Day Reconstruction Method (DRM). Manuscript under review.
- Robinson, J. P., & Godbey, P. (1997). *Time for life: The surprising ways Americans use their time*. University Park, PA: Pennsylvania State University Press.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-reports. *Psychological Bulletin*, 128, 934-960.
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication and questionnaire construction. *American Journal of Evaluation*, 22, 127-160.
- Schwarz, N., Sudman, S. (Eds.) *Autobiographical memory and the validity of retrospective reports*. (Springer Verlag: New York, NY, 1997).
- Stone, A. A., Shiffman, S. S., DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, N. Schwarz (eds.), *Well-being: The foundations of hedonic psychology* (pp. 61-84). New York: Russell-Sage.

A.2 Packet 3

How Did You Feel Yesterday?

Before we proceed, please look back at your diary pages.

How many episodes did you record for the Morning? _____

How many episodes did you record for the Afternoon? _____

How many episodes did you record for the Evening? _____

Now, we would like to learn in more detail about how you felt during those episodes. For each episode, there are several questions about what happened and how you felt. Please use the notes on your diary pages as often as you need to.

Please answer the questions for *every* episode you recorded, *beginning with the first episode in the Morning*. To make it easier to keep track, we will ask you to write down the number of the episode that is at the end of the line where you wrote about it in your diary. For example, the first episode of the Morning was number 1M, the third episode of the Afternoon was number 3A, the second episode of the Evening was number 2E, and so forth.

It is very important that we get to hear about all of the episodes you experienced yesterday, so please be sure to answer the questions for each episode you recorded. After you have answered the questions for all of your episodes, including the last episode of the day (just before you went to bed), you can go on to Packet 4.

First Morning Episode

Please look at your Diary and select the earliest episode you noted in the Morning.

When did this first episode begin and end (e.g., 7:30am)? Please try to remember the times as precisely as you can.

This is episode number _____, which began at _____ and ended at _____.

What were you doing? (please check all that apply)

commuting

working

shopping

preparing food

doing housework

taking care of your children

eating

- praying/worshipping/meditating
- socializing
- watching TV
- nap/resting
- computer/internet/email
- relaxing
- on the phone
- intimate relations
- exercising
- other (please specify _____)

Where were you?

- At home
- At work
- Somewhere else

Were you interacting with anyone (including on the phone, in a teleconference, etc)?

- no one → skip next question.

If you were interacting with someone (please check all that apply)

- spouse/significant other
- my children
- friends
- parents/relatives
- co-workers
- boss
- clients/customers/
- other people not listed _____

How did you feel during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

Appendix A. The Day Reconstruction Method (DRM)

Impatient for it to end	0	1	2	3	4	5	6
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/blue	0	1	2	3	4	5	6
Competent/capable	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Criticized/put down	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6

Appendix B

Experiment 7 Single Subject SCL

Analysis

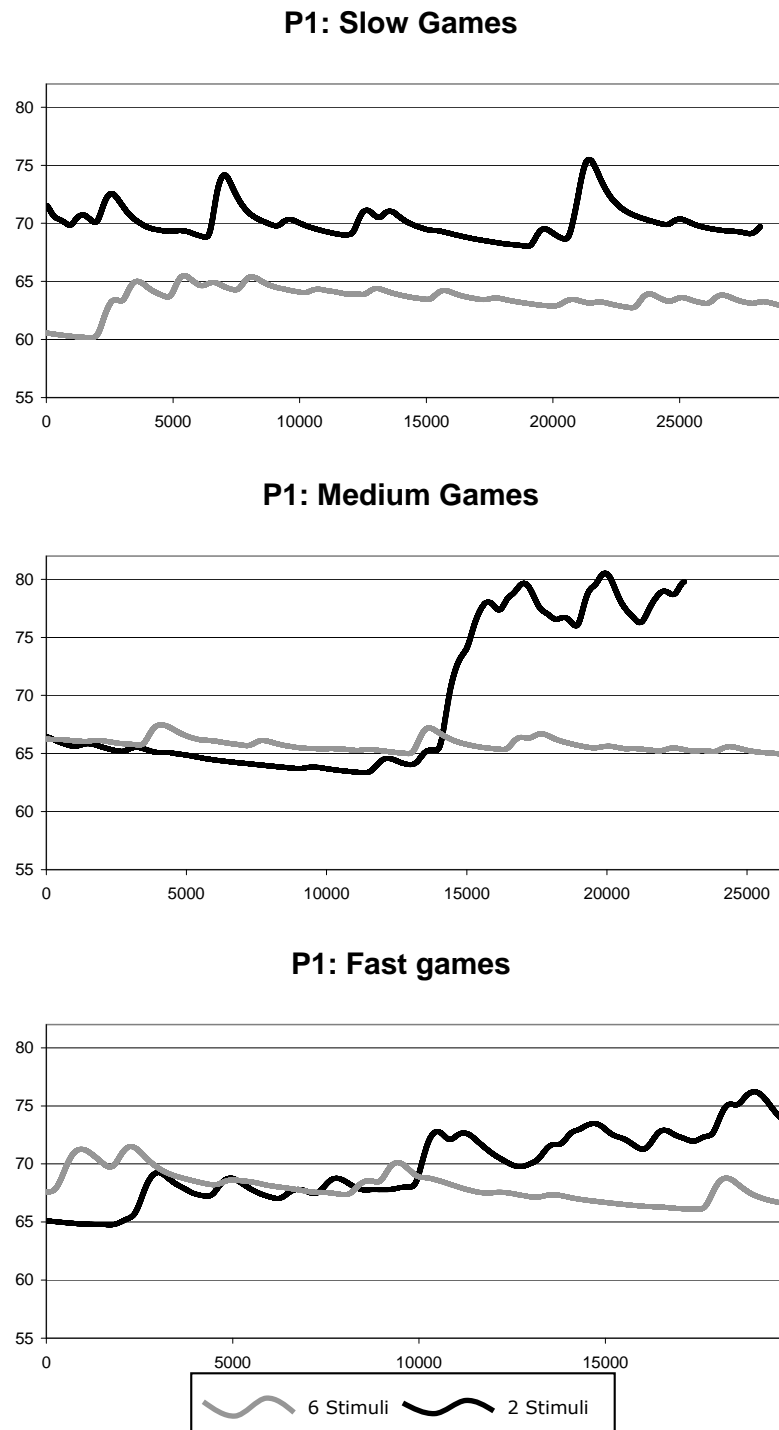


Figure B.1: SCL Data from all games played by Participant 1 in Experiment 4.1

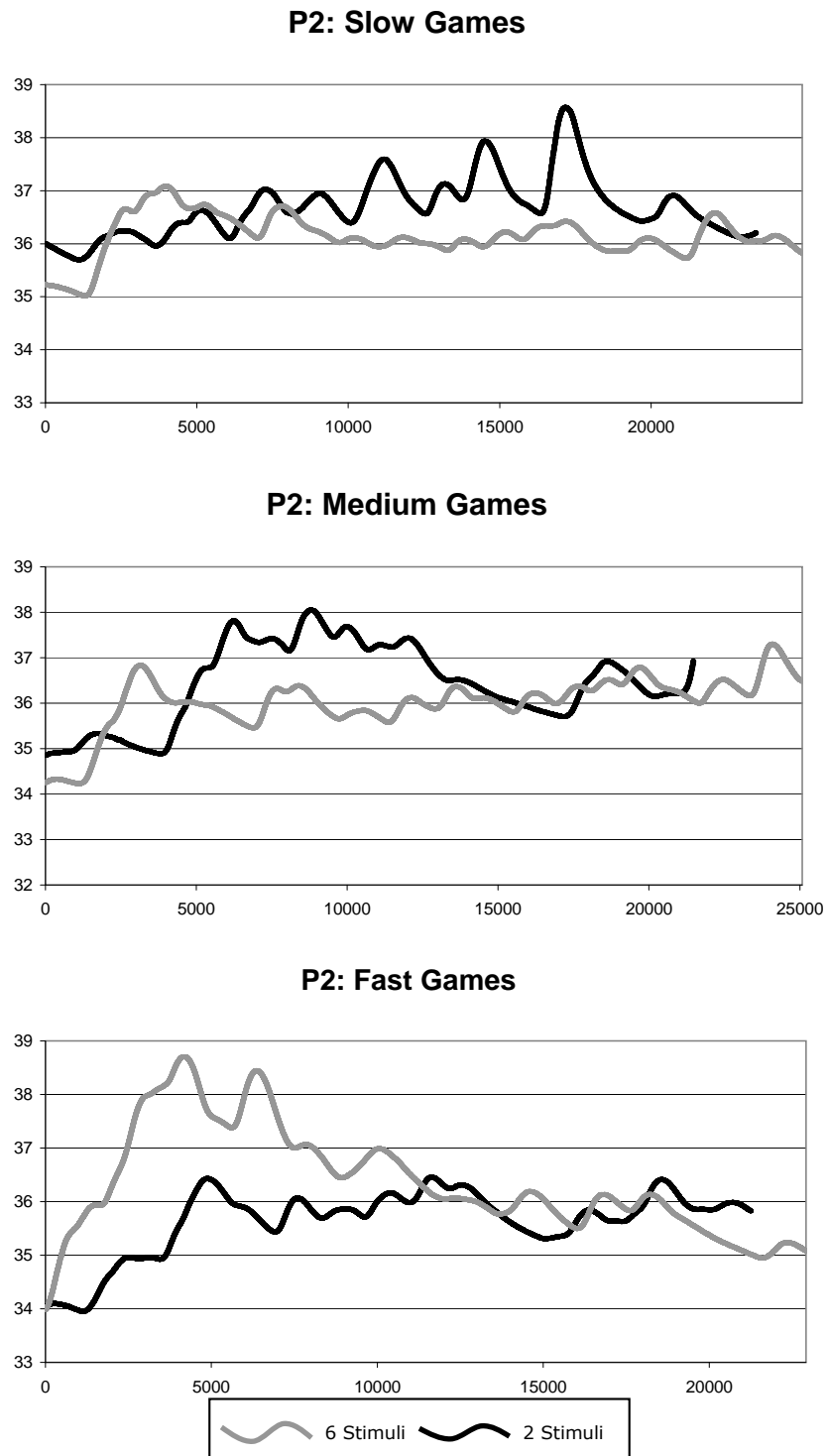
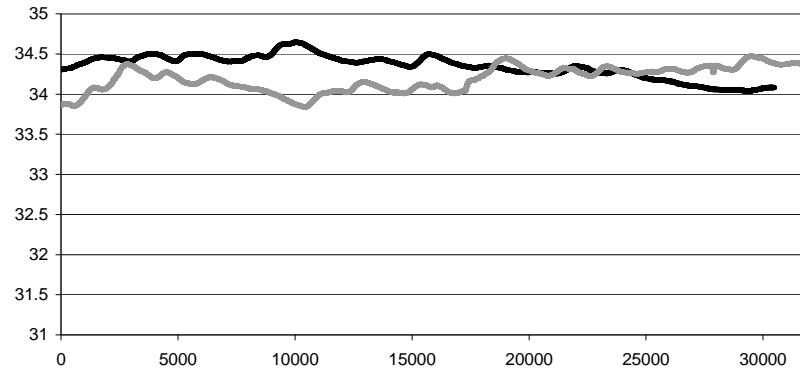
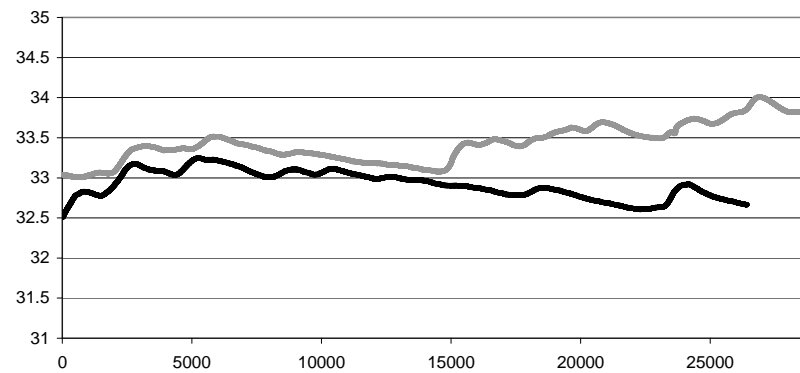


Figure B.2: SCL Data from all games played by Participant 2 in Experiment 4.1

P3: Slow Games



P3: Medium Games



P3: Fast Games

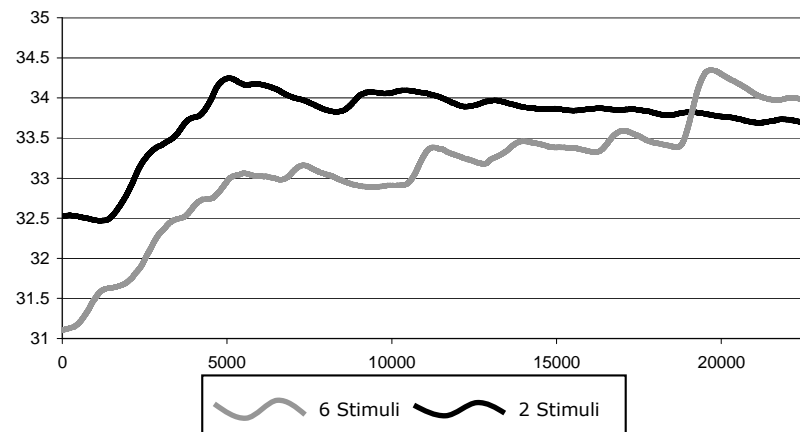
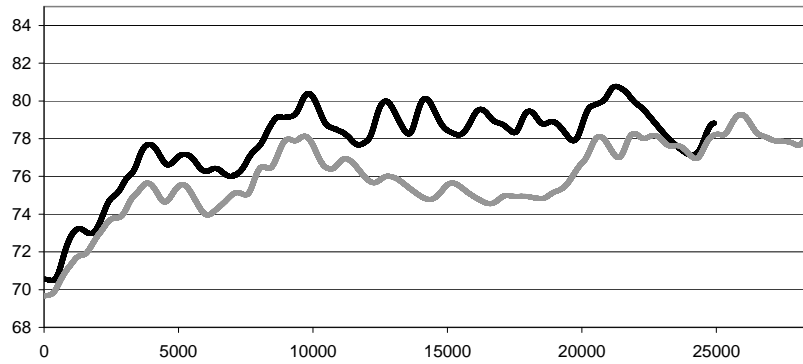
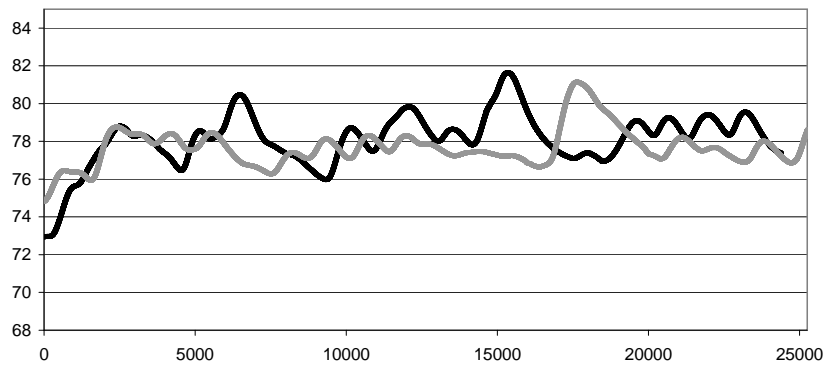


Figure B.3: SCL Data from all games played by Participant 3 in Experiment 4.1

P4: Slow Games



P4: Medium Games



P4: Fast Games

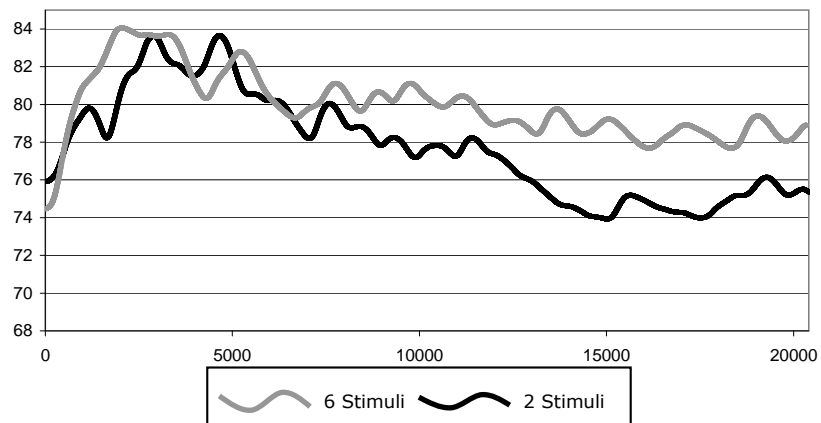
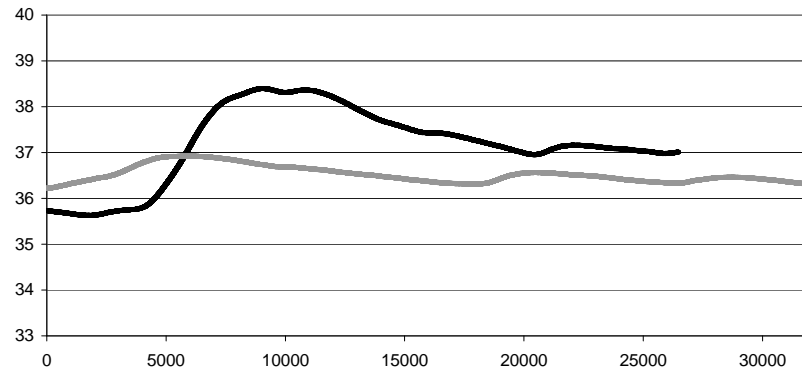
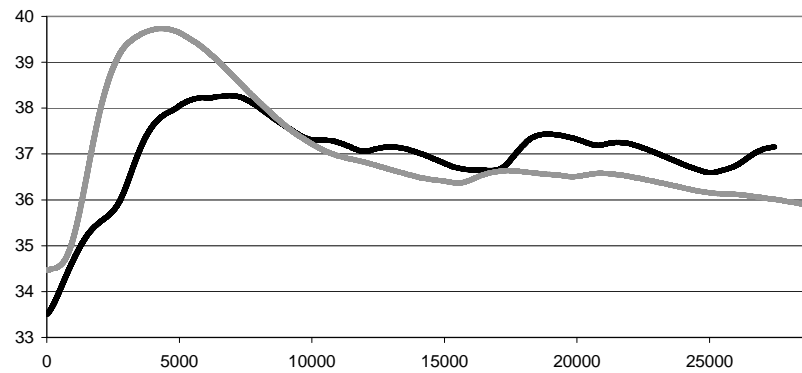


Figure B.4: SCL Data from all games played by Participant 4 in Experiment 4.1

P5: Slow Games



P5: Medium Games



P5: Fast Games

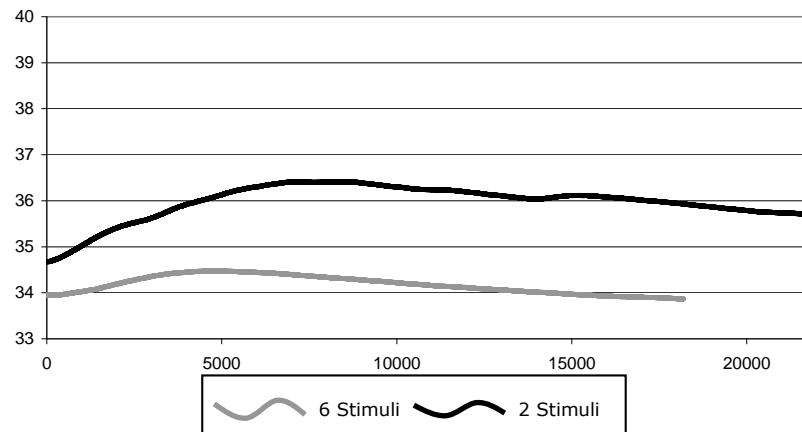


Figure B.5: SCL Data from all games played by Participant 5 in Experiment 4.1

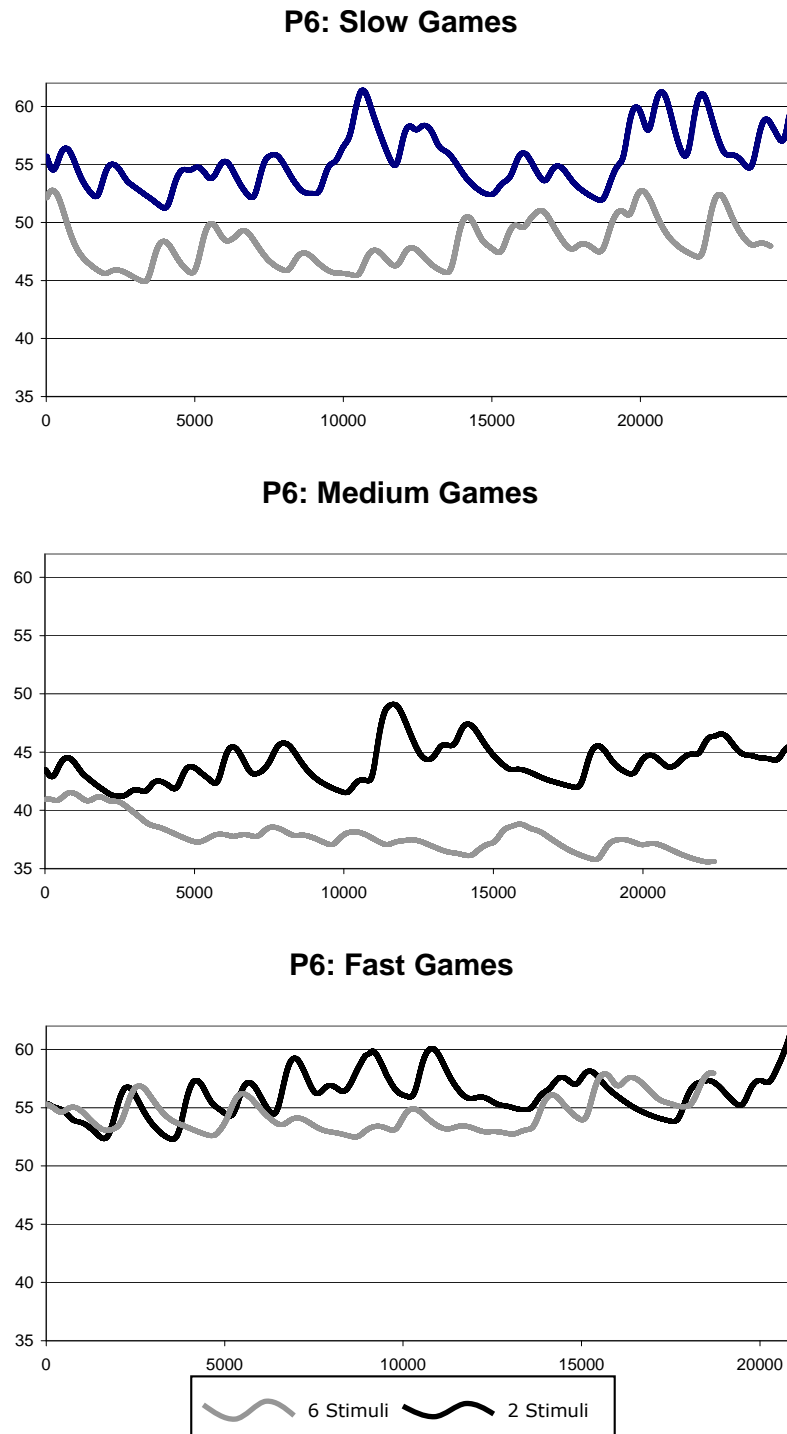


Figure B.6: SCL Data from all games played by Participant 6 in Experiment 4.1

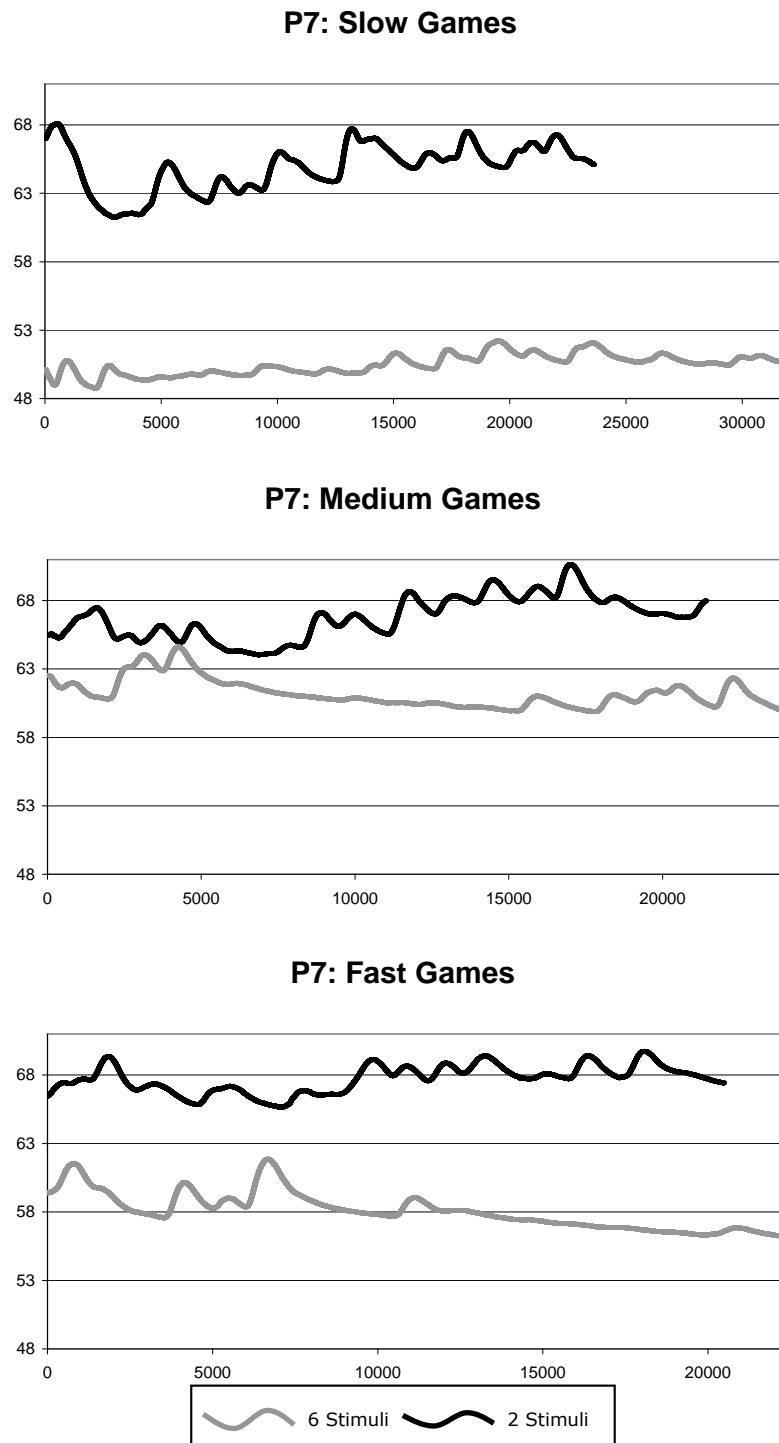


Figure B.7: SCL Data from all games played by Participant 7 in Experiment 4.1

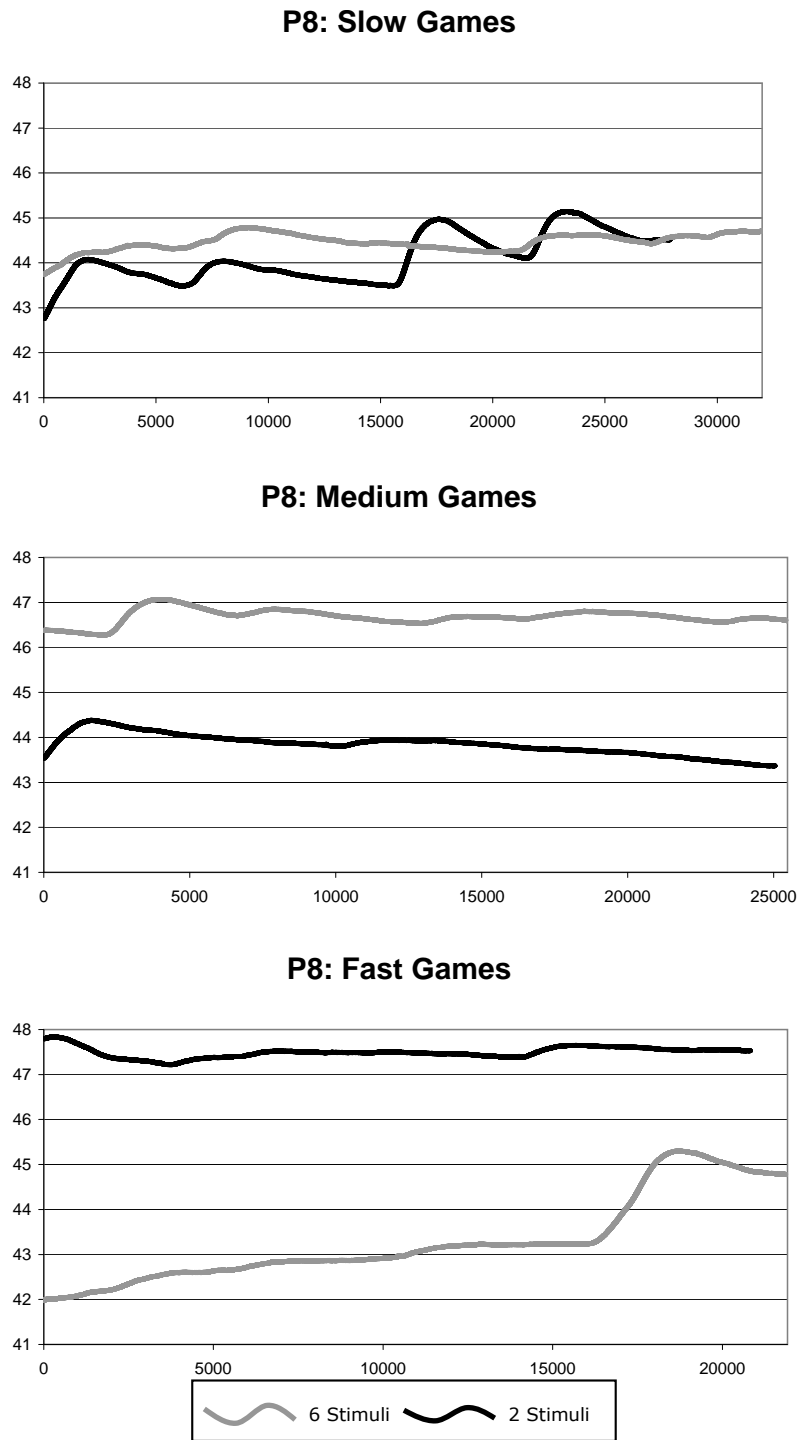
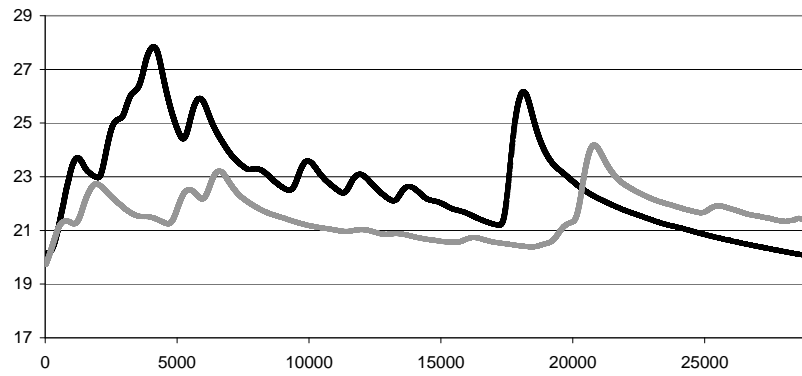
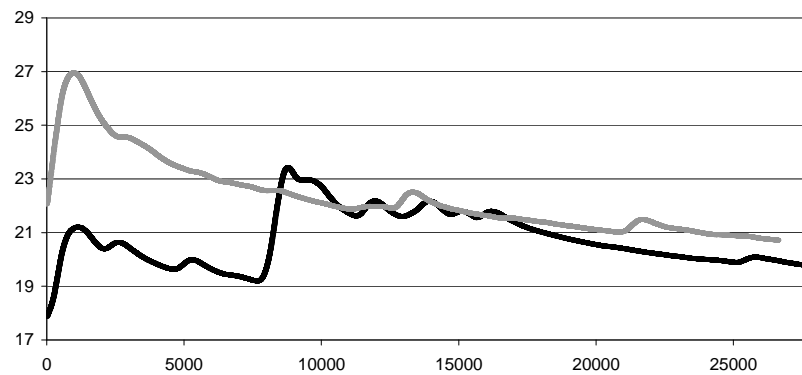


Figure B.8: SCL Data from all games played by Participant 8 in Experiment 4.1

P9: Slow Games



P9: Medium Games



P9: Fast Games

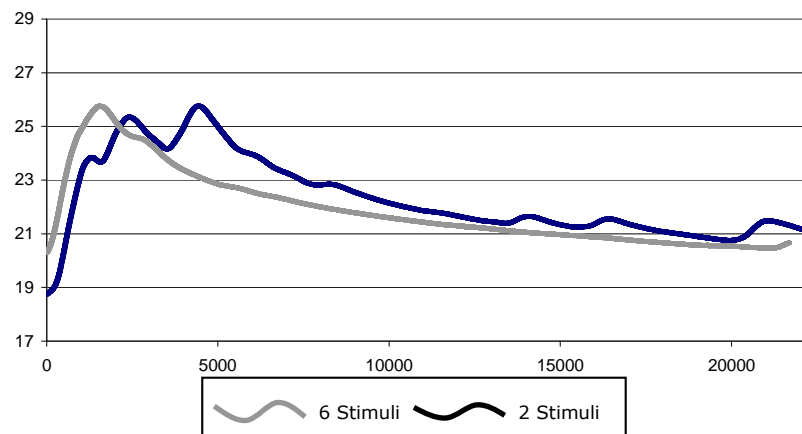


Figure B.9: SCL Data from all games played by Participant 9 in Experiment 4.1

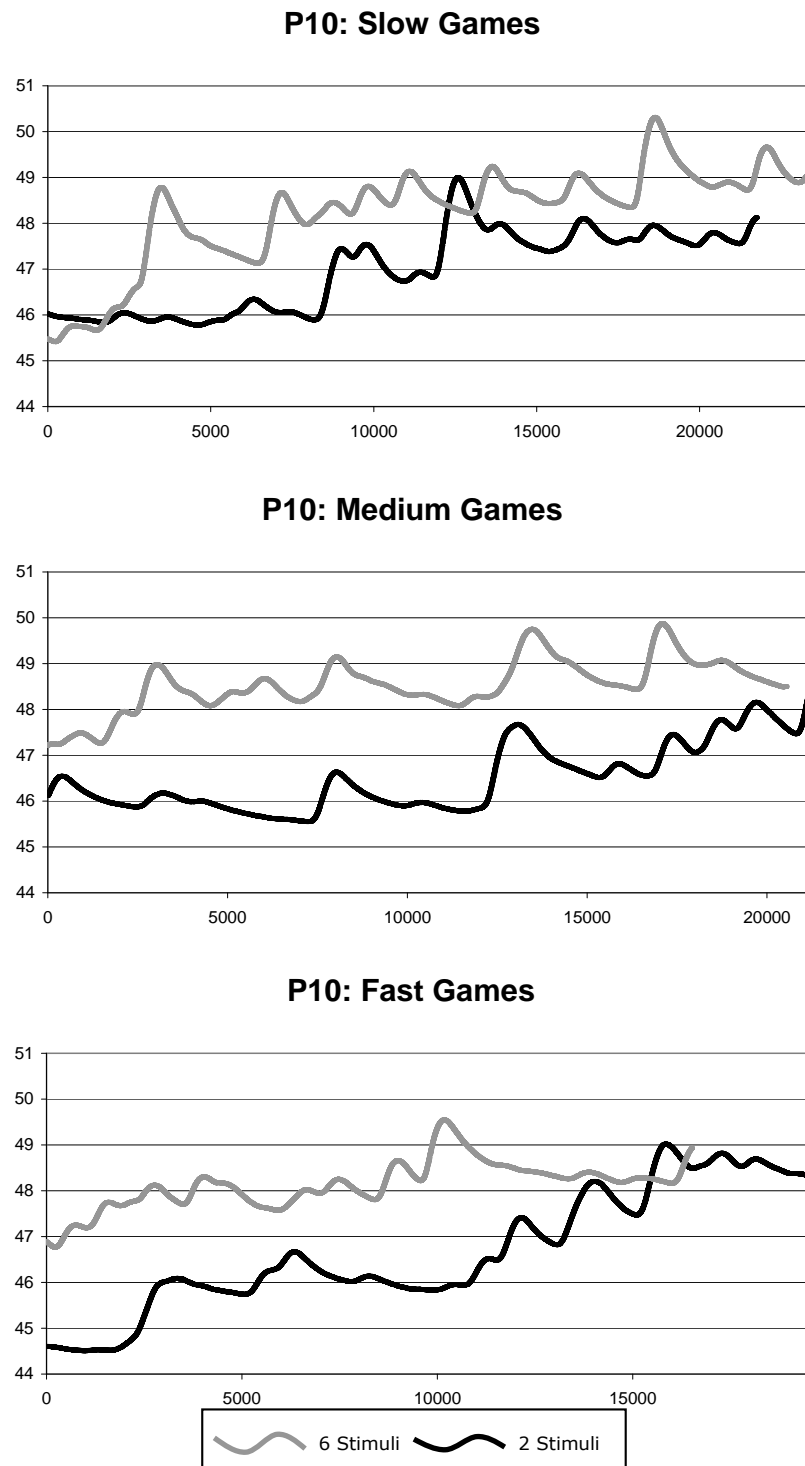


Figure B.10: SCL Data from all games played by Participant 10 in Experiment 4.1

Appendix C

Sequence of games played by each participant in Experiment 4.2

1. = 2-stimuli-slow
2. = 6-stimuli-slow
3. = 2-stimuli-medium
4. = 6-stimuli-medium
5. = 2-stimuli-fast
6. = 6-stimuli-fast

Table C.1: Sequence of games played by each participant in Experiment 4.2. Note that Stages 1 and 3 involved the presentation of all six games in randomised order, while the order of games presented in Stages 2 and 4 was determined by each participants' choice.

P	Stage 1						Stage 2						Stage 3						Stage 4					
1	6	5	1	3	2	4	5	3	1	2	4	6	5	2	6	1	3	4	4	5	4	4	4	4
2	1	5	4	6	2	3	5	2	4	3	1	6	3	1	5	4	6	2	4	6	6	6	6	6
3	4	3	1	6	2	5	3	5	1	6	4	2	4	3	1	6	2	5	3	4	4	3	3	3
4	5	1	6	4	3	2	5	3	4	1	2	6	5	2	6	1	3	4	5	4	4	5	6	2
5	5	2	6	1	3	4	4	1	6	2	5	3	3	1	5	4	6	2	4	6	5	3	1	2
6	1	2	3	4	5	6	5	4	3	6	2	1	4	2	3	1	5	6	5	3	3	4	4	4
7	1	5	4	6	2	3	6	1	2	5	3	4	2	6	4	5	1	3	6	6	4	6	5	5
8	3	1	5	4	6	2	3	1	5	2	4	6	5	2	6	1	3	4	2	2	2	2	2	4
9	1	5	4	6	2	3	3	5	1	2	4	6	3	2	6	4	5	1	3	2	4	3	4	4
10	4	3	1	6	2	5	2	4	6	5	3	1	4	2	3	1	5	6	4	1	2	4	2	2
11	1	2	3	4	5	6	4	5	6	3	1	2	5	1	6	4	3	2	5	4	6	6	5	6
12	5	1	6	4	3	2	2	5	4	6	3	1	6	5	1	3	2	4	2	4	4	5	1	3
13	4	2	3	1	5	6	1	3	2	4	5	6	3	2	6	4	5	1	4	2	5	4	6	3
15	2	6	4	6	1	5	1	2	6	5	4	3	2	3	4	6	1	5	1	4	2	3	4	4
16	5	5	6	4	3	2	6	4	5	3	2	1	6	5	4	3	2	1	6	6	6	6	6	6
17	4	6	3	1	5	6	4	5	3	6	2	1	4	3	1	6	2	5	6	4	5	6	4	2
18	2	2	4	5	1	3	4	6	5	3	1	2	1	2	3	4	5	6	1	2	4	6	X	X
19	6	2	1	3	2	4	5	3	4	6	1	2	6	5	4	3	2	1	5	5	6	2	4	1
20	2	5	4	5	1	3	4	5	2	1	3	6	5	2	6	1	3	4	2	3	5	6	1	4
21	4	3	3	1	5	6	3	1	5	2	4	6	3	1	5	4	6	2	3	1	4	1	3	5
22	5	3	6	1	3	4	1	3	5	2	4	6	5	2	6	1	3	4	4	6	5	1	2	5
23	6	2	4	3	2	1	5	3	2	4	6	1	6	5	1	3	2	4	4	6	6	6	6	6
24	4	3	1	6	2	5	6	4	2	5	1	3	4	3	1	6	2	5	5	6	2	5	4	4
25	4	3	1	6	2	5	3	1	4	5	2	6	4	2	3	1	5	6	2	4	4	5	3	6
27	1	2	3	4	5	6	1	3	6	2	4	5	1	2	3	4	5	6	1	2	3	5	1	3
28	4	3	1	6	2	5	1	3	5	4	2	6	2	6	4	5	1	3	1	1	1	3	5	5
29	2	3	4	6	1	5	3	2	6	5	4	1	4	2	3	1	5	6	2	5	5	4	6	1
30	3	1	5	4	6	2	5	6	2	1	3	4	1	5	4	6	2	3	1	3	5	5	5	5
31	6	5	1	3	2	4	2	5	3	4	6	1	6	5	1	3	2	4	5	1	4	3	2	4
32	2	6	4	5	1	3	1	3	5	2	6	4	3	2	6	4	5	1	1	3	2	4	6	1