The Impact of Avatar Fidelity on Social Interaction in Virtual Environments

Maia Garau

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London.

Department of Computer Science

University College London

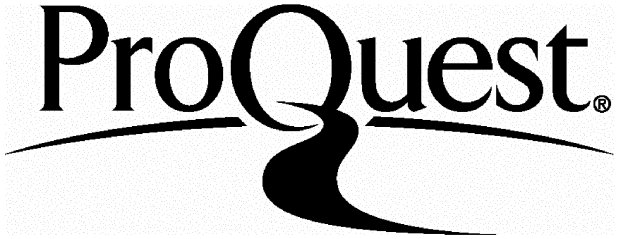
September 14, 2003

ProQuest Number: U643695

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.



ProQuest U643695

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my parents.

And to my Zia Paola, in loving memory.

Acknowledgements

During these years so many people helped me in so many ways, and I wish it were possible to thank them all. First and foremost I would like to thank my supervisor, Mel Slater, for giving me the freedom to find my own way and for always helping me towards clarity when I needed it most. My second supervisor, Angela Sasse, took me under her wing and always made me feel like an honourary member of her research group. An enormous thank you goes to Simon Bee, who helped me find my way and remained my mentor and friend even beyond his role as industrial supervisor. I would also like to thank my current industrial supervisor, Tim Stevens, for his encouragement and support throughout my stay at the labs.

The research described in this thesis comprises three large-scale experiments. A number of people collaborated on these with me, contributing their time and expertise. I would like to take this opportunity to thank them.

The first experiment on eye gaze conducted at BT Exact would not have been possible without Simon, who gave me precious guidance on the experiment and introduced me to the many helpful people who became my collaborators and friends. Special thanks go to Dickon Povey for helping me run the experiment over a period of several months. Alex Bourret helped set up the audiovisual link between the two soundproofed rooms, and Michelle Tasker helped me recruit participants; Paul Bowman, Jim Tasker, Mike Hollier, Dan Ballin and Dan Argent helped me find and set up equipment. Tim Child generously gave me use of Televirtual's avatars, and I am especially grateful to Sanja Abbott and Marcus Tutt for modifying their eye animations specifically for the purposes of my experiment. I would also particularly like to thank Nancy Pistrang for validating my grounded theory analysis of the interviews from this experiment.

For the experiment on responsiveness run at UCL, David-Paul Pertaub animated the avatars, worked with me on creating the library environment, and raised challenging questions to spur me along. Chris Barker and Nancy Pistrang offered many helpful suggestions on the design of the experiment and questionnaire. Michael Meehan and the NIH National Research Resource in Molecular Graphics and Microscopy at the UNC Chapel Hill kindly gave me use of their psychophysiological monitoring software. Special thanks go to Gillian Wilson for ad-

vising on the use of psychophysiological monitoring, Sharif Razzaque for advising on the setup and analysis, and Dirk Weirich for minding the monitoring equipment during the sessions.

For the experiment on eye gaze and photorealism run at UCL, Vinoba Vinayagamoothy was responsible for animating her own lower-realism avatar and for adapting the higher-realism avatars originally designed by David-Paul Pertaub. Special thanks go to Pip Bull and Anthony Steed for their precious help with the animations. Andrea Brogni created the virtual environment used in the experiment and helped me to run the experimental sessions.

For both the experiments run at UCL, I especially want to thank David Swapp for advising on general setup in the Cave, and Anthony Steed for his guidance on everything from DIVE to experimental design. The payment of participants was funded by the EQUATOR Interdisciplinary Research Collaboration. I would also like to gratefully acknowledge the participants for volunteering their time and thoughts.

Finally and most importantly, I want to thank my family and friends for their all their encouragement and good cheer throughout, and for keeping me happy during this time. A gratitude beyond words goes to my parents, who gave me the world, and to my aunt Paola, who filled it with wonder. This thesis is dedicated to them.

Abstract

The research presented in this thesis concerns the contribution of different levels of virtual human (or 'avatar') fidelity to social interaction in virtual environments (VEs). VEs present new possibilities for mediated communication by placing people in a shared 3D context. However, there are technical constraints in creating photorealistic and behaviourally realistic avatars capable of mimicking a person's actions or intentions in real time. At the same time, previous research findings indicate that virtual humans can elicit social responses even with minimal cues, suggesting that full realism may not be essential for effective social interaction. This research explores the lower boundaries of fidelity by investigating how different levels of responsiveness, photorealism and behavioural realism affect people's experience of interacting with virtual humans. The research presented comprises three between-group experiments involving over 200 participants. The experiments focus on distinct but interrelated questions.

In the first experiment, conducted in an immersive Cave-like system, participants explored a library containing a group of seated virtual readers. The aim was to investigate the degree to which the virtual humans were responded to as social entities as their responsiveness increased across the conditions. Results indicated that responsiveness significantly affected a range of responses, including the degree of sentience attributed to the agents. The findings signalled the need to further define the different social responses affected by virtual humans. Participants' heart rate and electrodermal activity were also recorded to explore the possibility of employing objective measures to study responses to virtual humans.

The remaining two experiments focused on behavioural realism and photorealism in the context of mediated communication between pairs of human participants. In the second experiment, participants interacted via a video-tunnel that provided them with a head-and-shoulders view of an avatar representing their conversation partner. Analysis of questionnaire responses indicated that the avatar whose gaze behaviour was inferred from conversational turntaking significantly outperformed the visually identical, but behaviourally less realistic, random-gaze avatar. An in-depth qualitative analysis of the interview responses yielded a theoretical model of the possible effects of avatar appearance and behaviour on perceptions of the communication

experience. For example, one outcome of the analysis with respect to gaze was the fact that excessive upward eye movement tended to undermine trust in the avatar. In the third and final experiment, participants were either in a Cave-like system or a head-mounted display. Each partner was represented to the other by a full-body and life-size avatar. The first goal of the experiment was to discover how the gaze models from the previous experiment would perform in a more demanding immersive setting; the second was to determine the impact of photorealism. Results indicated that the higher-realism avatar was significantly improved by the inferred-gaze model, whereas the lower-realism avatar was negatively affected. This suggested a strong interaction effect between appearance and behaviour. The findings from both experiments have implications for inexpensive ways of implementing avatar eye gaze for improved interaction in shared VEs.

The research presented in this thesis identifies a range of responses that can be affected by different levels of avatar fidelity. The overall conclusion is that even minimal behavioural cues can enhance the perceived quality of social interaction in VEs. However, the caveat is that there is a need for consistency between the fidelity of the avatar's behaviour and its appearance.

Contents

1 Introduction **21**

1.1 Research Problem 22

1.2 Research Questions 23

1.3 Contributions 25

1.4 Scope of the Thesis 25

1.5 Structure of the Thesis 26

2 Background **28**

2.1 VEs as a Communications Medium: Potential Benefits 29

2.1.1 First dimension: spatiality 31

2.1.1.1 Portrayal of directed attention 32

2.1.2 Second dimension: immersion 33

2.1.3 Third dimension: fidelity 34

2.1.4 Summary 35

2.2 State of the Art in CVEs 35

2.2.1 Online communities 36

2.2.2 Research on small group interaction in CVEs 38

2.3 Research Problem: Increasing Avatar Fidelity 42

2.3.1 Classifying virtual humans: avatars and agents 43

2.3.2 Goals for expressive avatars 44

2.3.3 The importance of nonverbal communication in face-to-face interaction: a brief review 45

2.3.3.1 Facial expression 45

2.3.3.2 Eye gaze 46

2.3.3.3 Body movements: gesture, posture and proxemics 47

2.4 Constraints on Avatar Fidelity 48

2.4.1 Constraints on visual fidelity: the tension between realism and real time 49

2.4.2 Constraints on behavioural fidelity: the tension between control and cognitive load 50
2.4.3 Prioritising aspects of avatar fidelity 53
2.4.4 Exploring the impact of minimal fidelity 56
2.4.5 Summary of research approach 59
2.5 Assessing Experiences in CVEs in terms of Presence, Social Presence and Co-presence 60
2.5.1 Presence 63
2.5.1.1 Definitions 63
2.5.1.2 Internal and external determinants 66
2.5.1.3 Measurement approaches 67
2.5.2 Short, Williams and Christie's original definition of social presence . . 72
2.5.3 Goffman's original definition of copresence 73
2.5.4 Appropriation of the original terms in the literature on CVEs 74
2.5.5 Appropriation of the term 'copresence' 76
2.5.6 Appropriation of the term 'social presence' 78
2.5.6.1 Social presence as the illusion of nonmediation 78
2.5.6.2 Social presence as access to another intelligence 78
2.5.6.3 Social presence as a subjective experience 79
2.5.6.4 Social presence as a multi-level response 80
2.5.7 Theoretical distinctions between social presence and copresence 81
2.5.8 Internal and external determinants 85
2.5.9 Measurement approaches 85
2.5.9.1 Summary: presence, social presence and copresence 88
2.6 Chapter Summary 88

3 Methodology **90**

3.1 Data Collection 91
3.1.1 Defining the research goals and expectations 91
3.1.2 Defining the independent and dependent variables 91
3.1.3 Experimental design 91
3.1.4 Piloting 93
3.1.5 Ethical considerations 93
3.1.6 Experimental procedure 94

3.1.7 Semi-structured interviews 95
3.1.7.1 Taping and transcription of semi-structured interviews 95
3.2 Data Analysis 96
3.2.1 Quantitative analysis of questionnaire data 96
3.2.2 Quantitative analysis of psychophysiological data 97
3.2.3 Qualitative analysis of interview data using grounded theory 98
3.2.3.1 Theoretical background: contextualising grounded theory within the quality/quantity debate 99
3.2.3.2 Grounded theory: guiding principles 101
3.2.3.3 Grounded theory procedures: the three stages of coding . . . 103
3.2.3.4 Tools used in grounded theory analysis 107
3.2.3.5 Quality control in grounded theory 108
3.2.3.6 Departure from the method 112
3.3 Chapter Summary 112
3.3.1 Reading guide to Chapters 4, 5 and 6 113

4 Experiment on Responsiveness **114**
4.1 Experimental Aims and Expectations 115
4.2 Experimental Design 115
4.2.1 Independent variables 115
4.2.2 Piloting 117
4.2.2.1 Piloting the task 117
4.2.2.2 Piloting the scenario 119
4.2.2.3 Piloting the pre-questionnaires 119
4.2.2.4 Piloting the post-test questionnaire 119
4.2.2.5 Piloting the agents 120
4.2.2.6 Piloting the procedure 120
4.2.3 Apparatus 120
4.2.3.1 ReaCTor 120
4.2.3.2 Psychophysiological monitoring equipment 121
4.2.4 Agents 121
4.2.4.1 Agent animations 121
4.2.4.2 Agent audio for the talking condition 122
4.2.5 Virtual environment 122

4.2.6 Task . . . 124
4.2.7 Population . . . 124
4.2.8 Procedure . . . 124
4.2.8.1 Phase 1: before entering the Cave . . . 125
4.2.8.2 Phase 2: inside the Cave . . . 125
4.2.8.3 Phase 3: after leaving the Cave . . . 125
4.2.9 Response variables . . . 126
4.2.9.1 Explanatory variables . . . 127
4.3 Results . . . 127
4.3.1 Copresence . . . 128
4.3.2 Presence . . . 129
4.3.3 Participant behaviour . . . 129
4.3.4 Perceived agent awareness . . . 129
4.3.5 Analysis of objective responses . . . 130
4.4 Chapter Summary . . . 131

5 Experiments on Eye Gaze and Photorealism 133
5.0.1 Research on the role and function of gaze . . . 134
5.1 Experiment on Eye Gaze . . . 135
5.1.1 Experimental aims and expectations . . . 135
5.1.2 Experimental design . . . 136
5.1.2.1 Independent variables . . . 136
5.1.2.2 Piloting . . . 137
5.1.2.3 Apparatus . . . 137
5.1.2.4 Avatars . . . 140
5.1.2.5 Avatar animations . . . 140
5.1.2.6 Task . . . 142
5.1.2.7 Population . . . 143
5.1.2.8 Procedure . . . 144
5.1.2.9 Response variables . . . 144
5.1.2.10 Explanatory variables . . . 146
5.1.3 Results . . . 146
5.1.4 Experiment summary . . . 148
5.2 Experiment on Eye Gaze and Photorealism . . . 149

5.2.1 Experimental aims and expectations 151
5.2.2 Experimental design 151
5.2.2.1 Independent variables 151
5.2.2.2 Piloting 151
5.2.2.3 Apparatus 152
5.2.2.4 Avatars 152
5.2.2.5 Virtual environment 155
5.2.2.6 Task 155
5.2.2.7 Population 155
5.2.2.8 Procedure 157
5.2.2.9 Response variables 157
5.2.2.10 Explanatory variables 159
5.2.3 Results 159
5.2.4 Experiment summary 164
5.3 Chapter Summary 164

6 Qualitative Analysis of Interview Data 166
6.0.1 Method of data collection 167
6.0.1.1 Semi-structured interviews 167
6.1 Impressions from Reading the Interview Transcripts 168
6.2 Goals of the Analysis 168
6.3 Method of Analysis 169
6.3.1 Phase 1: open and axial coding 169
6.3.2 Phase 2: axial and selective coding 171
6.3.3 Phase 3: selective coding 172
6.4 Analytical Findings 172
6.4.1 The key categories 172
6.4.2 The storyline 174
6.4.2.1 Communication expectations 174
6.4.2.2 Technical assumptions 175
6.4.2.3 Mirroring 176
6.4.2.4 Agency 177
6.4.2.5 Association 178
6.4.2.6 Personal contact 181

6.4.2.7 Spatial copresence 181
6.4.2.8 Impression formation 182
6.4.2.9 Trust 183
6.4.2.10 Core category: visual focus 184
6.4.3 Avatar fidelity: appearance and behaviour 184
6.4.3.1 Visual fidelity 185
6.4.3.2 Behavioural fidelity 186
6.4.3.3 Predictions 188
6.5 Chapter Summary 189

7 Discussion **192**
7.1 Experiment on Responsiveness 192
7.1.1 Copresence and presence 193
7.1.2 Responding to the agents as people 194
7.1.3 Attributed sentience 195
7.1.4 Objective responses 196
7.1.5 Summary discussion for experiment on responsiveness 197
7.2 Experiment on Eye Gaze 198
7.2.1 Summary discussion for experiment on eye gaze 199
7.3 Experiment on Eye Gaze and Photorealism 200
7.3.1 Perceived quality of communication 200
7.3.2 The impact of person variables 201
7.3.3 Additional measures 202
7.3.4 Summary discussion for experiment on eye gaze and photorealism . . . 204
7.4 Chapter Summary 204

8 Conclusions **205**
8.0.1 Experiment on responsiveness 206
8.0.2 Experiment on eye gaze 206
8.0.3 Experiment on eye gaze and photorealism 207
8.1 Key Questions Addressed in the Research 208
8.1.1 Main contributions 208
8.1.1.1 Methodological contributions 209
8.1.1.2 Substantive contributions 209
8.2 Directions for Future Work 212

A Summary of the three experiments **215**

B Materials for the Experiment on Responsiveness **217**

C Materials for the Experiment on Eye Gaze **228**

D Materials for the Experiment on Eye Gaze and Photorealism **234**

E Further Analysis for the Experiment on Gaze and Photorealism **253**

E.1 Additional response variables 253

E.2 Results 254

F Grounded Theory Analysis **256**

F.1 Sample Interview Transcript 257

F.2 Codes Generated in the Analysis 264

F.3 Process of Visual Abstraction 272

G Additional images from the experiment on responsiveness **276**

H Acronyms **278**

List of Figures

2.1 Comparison between VMC and IVEs along the dimensions of spatiality, immersion and fidelity 31

2.2 Using VMC, people remain in separate physical contexts and interact with each other via a video projection. Using IVEs, people interact in a shared, computer-generated 3D context where they are represented by digital proxies called 'avatars'. 34

2.3 Avatars used in the experiments by Slater, Tromp, Steed and colleagues 41

2.4 Technical constraints affecting avatar fidelity 49

2.5 The dimensions of visual fidelity include anthropomorphism, photorealism and truthfulness. 49

2.6 Aspects of virtual human fidelity considered in the research presented in this thesis 59

2.7 Classification of Physical presence, copresence and social presence, from IJsselsteijn, Freeman and de Ridder [IFdR01]. Italicised terms are added by Heeter [Hee01]. (Abbreviations: VR= virtual reality, LBE= location-based entertainment, SVEs= shared virtual environments, MUDs = multi-user dungeons.) . . . 82

3.1 Conditions set the stage for events. Context is shaped directly by causal conditions, mitigated by intervening conditions. 106

3.2 Components of the paradigm interrelate to link structure with process. Conditions set the stage for actions/interactions, resulting in consequences that can in turn shape new conditions. 106

3.3 Example of a mini-framework illustrating the intersection of two categories . . 108

3.4 Conceptual diagram illustrating emerging relationships between a selection of categories and their properties 108

3.5 Screen shot of an ATLAS.ti session. Concept labels appear on the right of the primary document. Primary documents, codes, memos and diagrams are all searchable, as are the interconnections between them. 109

4.1 Virtual 'readers' as seen upon entering the virtual library environment 115
4.2 Response zones for agents in conditions 3 *(Responsive)* and 4 *(Talking)* 116
4.3 Virtual readers in the library room would visually respond to the participant's proximity in the *responsive* and *talking* conditions. 117
4.4 Images showing participant in the virtual training room, then moving through the doorway to approach the characters in the virtual library 123
4.5 Means of raw questionnaire responses 128
4.6 EDA reading for random participant in condition 4 (*talking*). Dotted lines, indicating when an agent 'talks,' are followed by a sharp increase in EDA. 131

5.1 Both experimenters were able to follow the conversation on earphones while monitoring the participants and their corresponding avatars on the video wall. . 138
5.2 Video-tunnel setup 139
5.3 Video-tunnel setup showing the video condition 139
5.4 Diagram illustrating the video and audio output from one of the soundproofed rooms (CAB A) to the other room (CAB B) as well as to the videotape 140
5.5 The sensor was attached to the headphones. 141
5.6 Face-on view of male avatar looking 'at' and 'away' from partner 142
5.7 Face-on view of female avatar looking 'at' and 'away' from partner 142
5.8 Means of raw questionnaire responses 146
5.9 Participants in the Cave (left) and HMD (right). Participants in the HMD could not see their own bodies or physical surroundings while in the IVE; the image visible on the monitor was for the benefit of the experimenters. 152
5.10 Participants saw their conversation partner as a life-size avatar. The avatar was either lower-realism (left) or higher-realism (right) 153
5.11 From left to right: Lower-realism avatar, higher-realism female avatar, higher-realism male avatar 153
5.12 Two separate training rooms were separated from a central meeting room by doors that were only opened at the start of the conversation task. 155
5.13 Age range for participants 156
5.14 Means of raw questionnaire responses for lower-realism avatar 160

5.15 Means of raw questionnaire responses for higher-realism avatar 160

6.1 Conceptual network diagram illustrating substantive connections between categories 170
6.2 Conceptual network diagram illustrating the dimensions of a property 171
6.3 Conceptual network diagram illustrating a causal connection between two dimensionalised categories 171
6.4 Dimensions of 'mirroring' 177
6.5 Properties relating to the avatar's appearance 185
6.6 Properties relating to quality of behaviour 186
6.7 Properties relating to quantity of behaviour 187
6.8 Intersection between quantity (in terms of activity) and quality (in terms of precision) of animation 188

F.1 Detail of the first global diagram consolidating the individual conceptual network diagrams 272
F.2 Diagram focusing on the 'visual focus' category 273
F.3 Diagram focusing on the 'trust' category 273
F.4 Colour key 274
F.5 Diagram exploring the higher-level connections between key categories, now abstracted to coloured dots 274
F.6 Mini-diagrams illustrating the higher-level connections for each key category . 274
F.7 Diagrams summarising the location of key categories in the process model . . . 275

G.1 Participants were asked to attach the EKG sensors as illustrated, with the positive and ground leads on the left collarbone and the negative lead on the lower left rib. This arrangement is recommended by Meehan [Mee01] to maximise the amplitude of the signal while minimising artifacts. 276
G.2 The bipolar EDA sensors were attached as illustrated to the index and middle fingers of the participant's non-dominant hand. 276
G.3 The psychophysiological monitoring equipment was placed inside a wearable pack for the participant's comfort. This also ensured that the delicate optical cable remained protected, minimising signal failures. 277

G.4 Participants wore stereoscopic goggles and a head-tracking device, and held the 3D mouse in their dominant hand. The psychophysiological monitoring equipment was secured in a pack strapped to their waist. 277

List of Tables

2.1 Ellis, Gibbs and Rein's 'time space' groupware taxonomy, classifying interaction according to whether it is synchronous and co-located [EGR91] 30
2.2 Classification of CVEs 36
2.3 Comparison of three small group interaction experiments by Slater, Tromp, Steed and colleagues 39
2.4 Interfaces used in collaborative puzzle cube studies by Schroeder, Wideström and colleagues [SSA+01, WAS+00] 41
2.5 Summary of avatar requirements, after Benford et al. [BBF+95] 44
2.6 Approaches to driving avatar behaviours 51
2.7 The role of visual and temporal resolution in transmission of facial affect (after Ehrlich and Schiano [ES00]) 55
2.8 Differing views on the possible relationship between presence and social presence 62
2.9 Six conceptualisations of presence united by Lombard and Ditton under the banner of the "illusion of nonmediation" [LD97] 64
2.10 Subdivision of the term 'presence' into spatial and social aspects 66
2.11 Classification of possible external determinants of presence 68
2.12 Proposed measurement approaches 69
2.13 Original definitions of social presence and copresence reflecting complementary aspects of CVEs as a communications medium 75
2.14 Levels of social interaction proposed by Biocca and Harms, and Swinth and Blascovich 83
2.15 Conceptual distinctions between copresence and social presence 85
2.16 Summary of subjective measurement approaches for social presence 86
2.17 Summary of subjective measurement approaches for copresence 87

3.1 One-factor design with four levels 92
3.2 Two-by-two factorial design, each factor having two levels 92

3.3 Comparison between quantitative and qualitative approaches 100
3.4 Key differences between the hypothetico-deductive and grounded theory approach 102
3.5 Summary of main analytical tasks associated with each of the three phases of coding in grounded theory 103
3.6 Comparing the role of structure and process in Strauss and Corbin's paradigm . 105
3.7 Different approaches to grounded theory and their impact on evaluation criteria (after Willig [Wil01]) 111
3.8 Overview of the three experiments 113

4.1 Comparison of Hall's [Hal66] suggested measures of interpersonal distance and those adapted for use in the Cave 116
4.2 Probability and frequency of agent gaze at participant for each interpersonal zone 122
4.3 Factorial design for the experiment on responsiveness 124
4.4 Mean $\pm$ standard deviations of count response variables 128
4.5 Logistic regression with *perceived agent awareness* as the response variable . . 130
4.6 Mean $\pm$ standard deviations of heart rate response (beats per minute) 130
4.7 Mean $\pm$ standard deviations of EDA response (milliSiemens) 131

5.1 Summary of rules for 'while speaking' and 'while listening' animations 141
5.2 Factorial design for the first experiment on eye gaze 143
5.3 Mean $\pm$ standard deviations of count response variables 146
5.4 Logistic regression with *face-to-face* as the response variable 147
5.5 Logistic regression with *involvement* as the response variable 147
5.6 Logistic regression with *copresence* as the response variable 148
5.7 Logistic regression with *partner evaluation* as the response variable 148
5.8 Logistic regression for the sum of all response variables 148
5.9 Summary of key differences between the two experiments on eye gaze 150
5.10 The second experiment on gaze used head-tracking in all conditions, only varying the eye animations 154
5.11 Factorial design for the experiment on eye gaze and photorealism 156
5.12 Age distribution for participants 156
5.13 Mean $\pm$ standard deviations of count response variables for perceived quality of communication 161
5.14 Logistic regression for response variables for perceived quality of communication 161

5.15 Logistic regression for *social-copresence* and its subcomponents (*general copresence*, *spatial copresence* and *personal contact*) 162
5.16 Logistic regression with *gaze fidelity* and *avatar fidelity* as the response variables 163
5.17 Logistic regression with *presence* as the response variable 163

6.1 General themes identified prior to grounded theory analysis 168
6.2 Temporal overlap between the three phases of analysis 169
6.3 Key categories 173

A.1 Experimental design and running 215
A.2 Analysis and main findings 216

E.1 Logistic regression for the additional count response variables 254

Chapter 1

Introduction

Recent works of cyberfiction have depicted a not-so-distant future where the Internet has developed into a fully 3-dimensional and immersive datascape simultaneously accessible by millions of networked users. This virtual world is described as having spatial properties similar to the physical world and its virtual cities are populated by digital proxies of people, called avatars. Here people can interact with each other and with artificial intelligences (AIs) that are visually and sometimes behaviourally indistinguishable from humans. The multisensory sophistication of this shared space is such that it supports interpersonal communication on a level of richness interchangeable with face-to face interaction. Recent films such as *The Matrix* [Mat99] illustrated how enveloping and fully convincing such an experience might be.

This thesis explores the intersection between two distinct but increasingly interrelated research areas: collaborative virtual environments (CVEs) and mediated communication. The vision presented in fiction and film encapsulates two of the central goals not only of CVEs, but also of any communications medium. First, to enable groups of people to collaborate and interact socially in an efficient and enjoyable way, and second, to foster the illusion that people are together when in reality they are in distinct physical locations.

CVEs have the makings of a potentially powerful medium of communication that heralds new promises and challenges. It is the inherently spatial property of CVEs that sets them apart from other collaborative media. Though videoconferencing and groupware systems allow users to interact visually, the 3D context of each person's physical environment is lost. This can pose difficulties in small group interaction, where conversation management can be disrupted by ambiguous eye gaze cues. The loss of 3D context can also be particularly problematic in tasks for which it is essential to preserve spatial relationships, such as remote acting rehearsals. CVEs can begin to address these concerns by placing geographically dispersed users in a shared, computer-generated space where they can interact with the environment and with other users represented by avatars. Immersive interfaces can also offer multimodal, surround-

ing experiences that can create a strong sense of being inside that artificial space (presence), and sometimes of being there with others (copresence). As mediators of users' actions and appearance, avatars are likely to play a significant role in social interaction in CVEs.

Though not currently in mainstream commercial use, there are two distinct domains in which CVEs are currently being investigated: in online virtual communities and in experimental trials involving networked research laboratories. Since the mid-1990s, the increasing popularity of the Internet has been accompanied by the development of online communities where users can interact in shared environments using desktop PCs. These range from text-based MUDs (multi-user dungeons) to online graphical chat spaces that offer either text-based or audio interaction. Research in this area has explored such issues as online identity and community formation. Conversely, laboratory-based research experiments typically involve smaller groups of participants and highly specialised immersive interfaces including head mounted displays (HMDs) and immersive CAVE ^{TM}-like systems. Research in this area has typically focused on issues such as the impact of display type or the visual sophistication of the avatars on interaction.

1.1 Research Problem

One of the major drawbacks of CVEs is the relative paucity of avatar expressiveness compared with live human faces on video. Avatars in graphical chats vary widely in appearance and can exhibit lively behaviours, however they have been critiqued for serving merely as placeholders and failing to contribute meaningfully to conversation. The avatars used in collaborative laboratory-based studies are typically visually simplistic and have limited behavioural capabilities, such as the movement of a single arm for object manipulation. A significant challenge in developing CVEs as a communications medium is the development of expressive avatars capable of contributing to interaction.

Although CVEs can offer the benefits of spatial interaction and immersive experience, they remain low-fidelity compared with video-mediated communication (VMC); where VMC portrays objects and events from the real world, CVEs portray an artificial environment populated with artificial representations of people. In increasing avatar fidelity there are technical challenges as well as theoretical goals to consider. These affect both the avatar's static appearance (visual fidelity) and dynamic animation (behavioural fidelity).

In terms of the avatar's appearance, technical restrictions related to rendering and bandwidth mean that there is a tension between *realism* and *real time*. VE designers pay particular attention to exploiting the capacity of the human perceptual system to infer information from

limited but informative cues. Naturally it is not always admissible to take such shortcuts. The level of realism required depends on task requirements, for instance insufficient visual and haptic realism in a flight training simulation could result in disastrous consequences. It is arguable that communication is a more forgiving task in that it does not require full photorealism. The ability of humans to decode caricature and cartoons indicates that we do not require exhaustive photorealistic depictions to decipher the human form. Lessons from cartoon animation also indicate that photorealism is secondary to behaviour, provided that behaviour is convincing.

Achieving convincing avatar behaviour, however, introduces additional challenges. Social psychology research on face-to-face interaction has identified several nonverbal behaviours that serve a communicative function in the expression of emotion and in effective conversation management. These include facial expression, eye gaze, gesture, posture and proxemics (spatial behaviour). Body and facial tracking makes it possible to animate an avatar using motion data from a real person. Tracking equipment can, however, be expensive as well as intrusive for users. On a theoretical level, it is also questionable whether full tracking will be desirable in a medium that is prized for the control it offers over user embodiment. Being computer-generated, avatars afford control not only over appearance but also over behavioural expression, thereby potentially avoiding the pitfalls of nonverbal leakage that can occur in both face-to-face and video-mediated communication. However, full manual control over an avatar's actions would introduce unacceptable cognitive load. The problem of driving avatar behaviours that appropriately represent the users can therefore be summarised as the tension between *control* and *cognitive load.*

1.2 Research Questions

Given these technical constraints and theoretical considerations, the approach taken in the research presented in this thesis has been to explore the lower boundaries of avatar fidelity. The logic used by many VE designers is to exploit minimum cues to obtain maximum results. This research extends earlier studies by investigating whether minimal fidelity can contribute to social responses including the sense of copresence and positive perceptions of the interaction experience. It comprises three laboratory-based experiments addressing three nested questions. The first concerns the underlying premise that avatars can contribute to interaction experience at all:

Question 1: Can an avatar make a positive impact on perceptions of the interaction experience?

This question is addressed in the experiment presented in Section 5.1, where avatar-

mediated communication is compared with video-mediated and audio-only communication. Immediately following from this is the related question concerning the contribution of minimal behaviours:

Question 2: Can minimal behavioural fidelity offer an improvement over mere 'liveliness'?

This question is addressed in the same experiment, by comparing a random-gaze avatar with an inferred-gaze avatar whose head movements are tracked and whose eye movements reflect conversational turntaking. Although the behaviours of the inferred-gaze avatar are not an exact reflection of the behaviours of the person represented, they nevertheless represent an increase in fidelity. Behavioural fidelity is therefore 'minimal' in two ways: firstly, a single behaviour is isolated for investigation (eye gaze). Secondly, the eye animations are inferred, rather than being a direct representation of a real person's eye movements in real time. A second experiment, reported in Chapter 4, investigates another key aspect of behavioural fidelity by comparing incremental increases in responsiveness. It addresses the question of whether behavioural fidelity can impact on copresence in the absence of verbal interaction.

The third question follows from the second, and addresses the assumption made by numerous researchers that behavioural fidelity should be prioritised over visual fidelity in the development of expressive avatars:

Question 3: What is the relationship between visual and behavioural fidelity?

This question is addressed in the third experiment, presented in Section 5.2. This experiment extends the earlier experiment on eye gaze, implementing the random-gaze and inferred-gaze models in avatars with different levels of photorealism. It addresses the question of whether improvements in behavioural fidelity benefit all humanoid avatars regardless of their appearance, or whether appearance tempers the impact of behaviours.

An additional question concerns the research methods employed to study people's sense of being with others in a shared VE:

Question 4: How should copresence be defined and measured?

In the growing body of research dedicated to social interaction in CVEs, the terms *copresence* and *social presence* are often used interchangeably to describe this sense of being together. Chapter 2 addresses the lack of consensus on the definition and measurement of these constructs. Chapter 3 discusses the use of objective psychophysiological data, as well as the combination of post-test questionnaires and qualitative analysis of interview data using grounded theory.

1.3 Contributions

This thesis makes both substantive and methodological contributions to the study of social interaction in CVEs. The main contributions are listed below.

Methodological contributions:

1. A proposed approach for adapting the grounded theory method to conduct an in-depth qualitative analysis of interviews with participants (Chapter 3). Grounded theory analysis is particularly suited to eliciting causal models of user responses to relatively new phenomena and can be combined with questionnaire-based approaches.

2. A negotiation task for use in studies on dyadic communication (Chapter 5).

Substantive contributions

1. A detailed critical review of the problematic use of the terms social presence and copresence in the current literature on CVEs (Chapter 2).

2. Research findings that address the impact of different aspects of avatar fidelity on social responses to avatars (Chapters 4-6). These findings also have implications for inexpensive approaches to increasing avatar expressiveness.

1.4 Scope of the Thesis

The appearance of avatars can range from abstract to animal to humanoid, and from cartoonish to photorealistic. This thesis is concerned exclusively with humanoid avatars. Its main focus is on subjective responses to varying levels of visual and behavioural fidelity, although one experiment makes exploratory use of psychophysiological measures to study objective responses. The challenge of driving appropriate avatar behaviours partially shaped the incentive to study the lower boundaries of behavioural fidelity. This research is not concerned, however, with the techniques involved in creating or animating avatars, but with their perceived contribution to experiences of interaction.

The purpose of the three experiments presented in this thesis was to isolate the impact of avatar fidelity as far as possible from potentially confounding factors such as real-life relationships, explicit technical knowledge of how the avatars were animated, or communicative conventions derived from long-term use. The research therefore does not directly address the issue of self-presence, or of participants' relationship with their own avatar.

1.5 Structure of the Thesis

Chapters 2 and 3 are introductory and cover relevant research and methods. Chapters 4 to 6 present the design and findings of three studies investigating responses to avatars. Chapters 7 and 8 draw conclusions from the findings and propose directions for continuing research.

Chapter 2 contextualises the research by discussing its motivation, the central problem addressed, and the general approach taken. Although CVEs potentially support spatial and fully immersive interaction, one significant barrier to interaction is the limited expressive potential of avatars. The technical challenges involved in increasing avatar expressiveness shaped the general approach taken, namely to explore the potential contribution of minimal fidelity to social responses. One crucial response, copresence, is discussed in detail in relation to presence and social presence.

Chapter 3 focuses on the methods used to address the research questions. It introduces features of experimental design common to all three experiments, as well as the method of statistical analysis used to analyse the questionnaire data. It also discusses the use of heart rate and electrodermal activity to measure objective responses in one of the experiments. Finally, it addresses the combination of qualitative and quantitative methods of analysis with particular reference to grounded theory. Grounded theory is a qualitative research method used to elicit detailed causal models of participants' responses to avatars from interview data.

Chapter 4 presents findings from an experiment designed to investigate social responses to avatars in the absence of verbal interaction. The experiment compares the impact of visually identical virtual humans that differ only in their reactive behaviours towards participants. The responses considered include subjective questionnaire data, as well as objective psychophysiological data.

Chapter 5 presents two closely related experiments on eye gaze. Section 5.1 presents an experiment on the impact of eye gaze on dyadic communication in a non-immersive setting. Eye gaze is of central importance in face-to-face communication, since it serves both to communicate attention and to manage the smooth flow of conversation. An avatar with random gaze is compared with a visually identical one with 'inferred' eye gaze. The avatars are compared to audio-only and full video conditions, with an aim to understanding how they perform in relation to other media when they offer minimal behavioural cues. The experiment presented in Section 5.2 directly extends the previous study, investigating the impact of inferred and random gaze in an immersive setting. It also investigates the impact of the animations on lower-realism and higher-realism avatars in order to explore the relationship between visual and behavioural realism.

Chapter 6 discusses the grounded theory analysis of the interviews from the first experiment on eye gaze. It expands on questionnaire results presented in Section 5.1, focusing on aspects of the experience deemed important by respondents. The analysis highlights ways in which expectations and context can shape the way aspects of avatar fidelity are interpreted by participants, and how these interpretations can affect the role played by the avatar in the communication process.

Chapter 7 discusses the overall findings and the implications that can be drawn from them. Finally, *Chapter 8* draws conclusions and gives suggestions for future work.

Chapter 2

Background

There are numerous application areas for virtual environments (VEs), from simulation to training to the treatment of phobias. This thesis focuses on CVEs[1] as an emerging 3D communications medium. Although their use is not as yet widespread, it is likely that as the technology matures they will provide increasingly rich possibilities for interaction. In the words of Schroeder, "We don't know yet how the technology will mature: walk-in, Cave-type, large-screen, HMD, more sophisticated desktop CVEs, or some other. It is a good bet, however, that *some* kind of technology for interacting with others inside CVEs, with graphics that provide a sense of being there and high quality audio, *will* become widespread" [Sch02a, p.291].

This chapter contextualises the research by discussing relevant literature that has shaped the driving motivation, the problem addressed, and the approach taken in this thesis. It is divided into four main sections. The first explores the potential strengths of CVEs as a communications medium. The second discusses the current state-of-the-art in CVEs. Though not yet in mainstream use for communication purposes, CVEs are primarily studied in two areas: in online large-scale virtual communities, and in research institutions. Though distinct in aims and scope, research in both areas highlights similar problems and limitations in terms of the effectiveness of graphical embodiments for social interaction and collaboration. The third section contextualises the research problem by discussing the goals in creating expressive avatars for communication purposes, and some technical constraints on the level of visual and behavioural fidelity achievable in current CVEs. Given these constraints, it argues for an exploration of how minimal fidelity might be exploited to achieve maximum results in terms of people's sense of being in the shared virtual space together with others. The fourth and final section addresses the challenges of defining and measuring this sense in terms of presence, social presence and copresence.

[1] CVEs are sometimes also referred to as shared virtual environments (SVEs)

2.1 VEs as a Communications Medium: Potential Benefits

Anthony Giddens described face-to-face talk as a communications medium [Gid84]. In this thesis, the term 'communications medium' applies exclusively to interpersonal, mediated interaction between geographically dispersed people, or between people and artificial social entities. One of the underlying assumptions behind research in both video-mediated communication (VMC) and CVEs has been that the inclusion of visual information can improve mediated interaction by harnessing our natural ability to read meaning into the human form. Short, Williams and Christie have argued that all attempts at producing visual communications media are "primarily directed at remedying what is the most obvious defect of the simple telephone–the fact that one cannot see the other person or group" [SWC76, p.43]. The question that arises with the advent of CVEs is, what happens when both the environment and the people in it are not portrayals of the real world, but artificial simulations?

CVEs are networked, computer-generated environments capable of supporting human-to-human communication by allowing users to interact with the space and with each other via graphical embodiments called avatars. This thesis employs the term 'collaborative' in the broadest sense, as "any activity involving a series of tasks within a virtual environment that requires social and cooperative efforts between users within a group" [CW00, p.183]. In this definition, CVEs include not only environments used explicitly for work-related purposes, but also for social interaction and play. CVE applications can range from conferencing, simulation and training, shared visualisation and collaborative design, to social communities and multiplayer games. Avatars play a significant role in all of these contexts because they embody the user in a shared space, opening multiple possibilities for interaction.

CVE research is cross-disciplinary, drawing from fields including computer science, psychology, sociology, architecture, urban planning and human-computer interaction. The study of collaboration in CVEs relates closely to the field of computer-supported cooperative work (CSCW), which is concerned with investigating how computers can facilitate human interaction. CSCW technology is commonly referred to as groupware, defined by Ellis, Gibbs and Rein as "computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment" [EGR91, p.40]. Groupware systems differ from single-user applications in that they reflect the activities of multiple users in the environment, therefore actively supporting group communication, collaboration and coordination.

A number of technologies can be classified as groupware, and the authors propose a taxonomy based on the temporal and spatial properties of interaction (Table 2.1). This taxonomy

accounts only for 2D media. The category of 'synchronous distributed interaction' includes computer conferencing technologies that combine different configurations of document sharing facilities and live video of participants. Though not explicitly included in the taxonomy, CVEs also belong to this category; like videoconferencing, they differ crucially from face-to-face interaction in that communication is synchronous, but participants occupy distinct physical spaces.

Table 2.1: Ellis, Gibbs and Rein's 'time space' groupware taxonomy, classifying interaction according to whether it is synchronous and co-located [EGR91]

	Same time	**Different times**
Same place	Face-to-face interaction (e.g. meeting room technology)	Asynchronous interaction (e.g. physical bulletin board)
Different place	Synchronous distributed interaction (e.g. desktop conferencing)	Asynchronous distributed interaction (e.g. E-mail)

Benford et al. argue that there has been an increased interest in spatial approaches to CSCW, reflecting a shift in focus towards the *context* as opposed to the mere *process* of collaboration [BBRG96]. They classify CSCW systems that emphasise spatial interaction according to three dimensions: *spatiality*, *transportation* and *artificiality*. For the sake of consistency with discussions on fidelity and presence later in this chapter, these categories will be referred to as follows:

1. **spatiality**: the degree to which participants are provided with a shared and navigable spatial context;
2. **immersion**: the degree to which participants are provided with a surrounding sensory experience, resulting in a sense of 'transportation' from their physical surroundings to the mediated context;
3. **fidelity**: the degree to which sensory information in the mediated context is based on information from the real world.

While it is not the aim of this section to compare the relative merits of video and avatar-mediated communication, the discussion of some key properties along these three dimensions highlights some of the potential strengths of CVEs as a medium. Figure 2.1 illustrates some distinctions between videoconferencing and CVEs. Videoconferencing portrays participants' real appearance and actions and is therefore high in fidelity; however, it is experienced on a 2D screen and is therefore low in spatiality and immersiveness. Conversely, immersive VEs

(IVEs) provide a 3D surrounding experience and are high in spatiality and immersiveness, but low in fidelity because portrayals of participants and the environment are synthetic. The general aim of CVE research is to increase fidelity with a view to bridging the gap between virtual and face-to-face interaction. The remainder of this section will discuss these three dimensions in more detail.

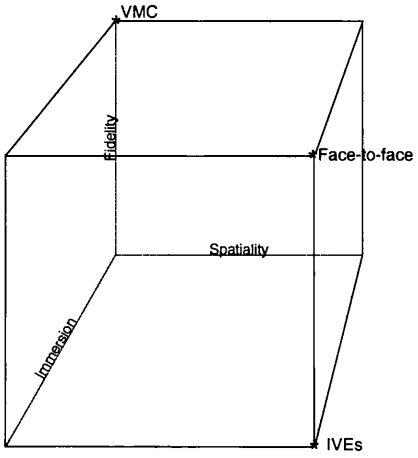


Figure 2.1: Comparison between VMC and IVEs along the dimensions of spatiality, immersion and fidelity

2.1.1 First dimension: spatiality

Physical space plays a central role in human interaction. Harrison and Dourish have argued that by being invested with social significance, a 'space' becomes a 'place' that shapes specific social conventions [HD96]; for instance, though concert halls and rooms in conference centres may both contain a stage, expectations for 'performer' and 'audience' behaviours are very different. Space also plays a significant role in regulating ongoing conversation and the formation of groups through the adjustment of interpersonal distance [AD65] and orientation [SS72] (see Section 2.3.3.3).

Systems emphasising spatial aspects of interaction include media spaces, spatially oriented video and CVEs. Media spaces represent an attempt to bridge separate physical spaces through a permanently active electronic connection. The Xerox PARC media space is an example of a system designed to enhance a sense of 'embeddedness' between two geographically separate working groups [HBAM97] by maintaining a constant live video feed between the two sites. Spatially oriented video systems such as Hydra [Sel95] attempt to preserve spatial cues among participants in order to accurately represent directed attention in group interaction.

It is their inherent spatiality that sets CVEs apart from other groupware systems. Benford,

Dourish and Rodden argue that "CVEs are perhaps the most radical form of spatially oriented cooperative systems. Emerging from research into real-time graphics, virtual reality, and computationally shared workspaces, CVEs are virtual environments which can be simultaneously occupied by distributed individuals, who come together to achieve some collective activity as much as, in the everyday world, we gather to work in meeting rooms, in offices, in hallways, or around the coffee machine" [BDR00, p.1]. Being in a shared space enables people to share artifacts and maintain peripheral awareness of surrounding activities [BGRP01].

Though media spaces enable people to share visual information from their physical environment [HFH+98], the disadvantage is that the 3D context of each user's physical environment is lost [TR00]. While there are many advantages to using different video configurations for remote communication, there are certain collaborative situations, such as remote acting rehearsals, in which it is essential to preserve spatial relationships. CVEs begin to address these needs by placing users in a shared 3D context. Experiments with virtual acting have shown that even through desktop interaction with very limited interfaces, actors are able to build a shared notion of the space, allowing them to accelerate the important 'blocking' phase of rehearsal where they plan their movements in relation to actors and props on stage [SHS+00].

Benford et al. argue that users' relationship to space distinguishes CVEs from other shared information groupware systems: "Like many applications, the general model of shared information systems places users externally to the space; they are observers looking at information through a window (the screen). Users are not generally embodied within these systems and tend not to be directly visible to one another (or if they are, it is in a very limited way such as through a telepointer). In contrast, within the spatial model users inhabit the shared space and are directly embodied" [BBF+94, p.666]. It is this characteristic of user embodiment that helps in addressing the challenge of portraying directed attention.

2.1.1.1 Portrayal of directed attention

One of the central problems in mediated communication is the portrayal of directed attention. The most commonly used multiparty videoconferencing setup is the picture-in-picture (PIP) approach, where each participant's monitor contains subscreens showing live video feed from other participants. Subscreens are scaled down to accommodate increasing numbers of participants, making the system limited in its scalability. The other significant drawback is that the lack of spatial cues and camera offset make it difficult to determine where a participant's attention is focused [Whi95]. Buxton and colleagues have pointed out that the PIP approach fails to support directed gaze, disrupting turntaking in group conversation [Bux92]. Short, Williams and Christie argue that with camera offset "the regulatory function of eye-contact may thus be

worse than removed, its operation may be reversed" [SWC76, p.55].

A number of groupware systems have been designed to address this shortcoming. These include the videotunnel, designed to preserve gaze in dyadic interaction [Bux92], and the Hydra system, designed for spatially consistent gaze in multiparty interaction [Sel95]. Other systems such as MAJIC [OMIM94] and Clearboard [IKG93] also include a shared workspace, so that participants can monitor each other's directed attention on the common task.

The advantage of CVEs is that participants' embodiments can be seen in spatial relation to each other and to the objects they are interacting with. According to Hindmarsh et al., this means that "unlike media spaces, participants are visibly 'embodied in' and 'connected to', the common world of objects" [HFH+98, p.218]. Unlike videoconferencing and media spaces where camera positions are fixed, participants in CVEs are free to control their point of view (POV) by navigating through the environment. As Bowers, Pycock and O'Brien point out, this alone allows a degree of awareness of other's focus of attention [BPO96]. However, the granularity of this understanding depends largely on the fidelity of the embodiment, on its level of visual detail (photorealism) and behavioural accuracy. Fraser et al. report that even simple humanoid avatars in their system raised expectations for a 'humanlike' field of view (FOV) that was not supported by the system, leading to incorrect assumptions about what others could see [FGV+00]. There are significant challenges involved in portraying accurate eye gaze in CVEs, particularly in an immersive setting where participants' faces are partially obscured by stereoscopic goggles, making tracking more problematic.

In summary, the portrayal of space in CVEs has two practical advantages for remote collaboration: the provision of a shared interaction context for geographically dispersed users, and the portrayal of directed attention. Though avatars in CVEs offer some potential advantages in terms of spatial position and orientation, accurate gaze portrayal represents a nontrivial problem.

2.1.2 Second dimension: immersion

VEs can be experienced non-immersively using a desktop, or immersively using an HMD or CAVE ™-like system[2]. Desktop VEs can suffer from the same limitations in field of view as videoconferences. In IVEs, however, stereoscopic images, combined with head-tracking, produce "the effect of a circumambience, or a 'virtual world' that surrounds the user. This circumambience creates the feeling that you are immersed in a computer simulated environ-

[2]CAVE ™ is a trademark of the University of Illinois at Chicago. In the remainder of this thesis the term 'Cave' will be used to describe the generic technology as described in [CNSD93] rather than to the specific commercial product.

ment, that is, the feeling that you are experiencing the computer simulation from the inside, not merely passively observing it from the outside" [BW93, p.699]. IVEs offer not only a visually surrounding environment, but also a multisensory experience. In addition to the audiovisual interaction possible through other media, they also open the possibility of haptic interaction. Durlach and Slater posit that the haptic modality may have the greatest potential for enhancing the sense of 'togetherness' in VEs [DS00].

Swinth and Blascovich argue that the immersive property of IVEs broadens the communicative bandwidth in comparison with other media: "Because of its immersive, interactive nature and its ability to render both auditory and visual signals (and perhaps haptic and olfactory signals in the future), a greater variety of verbal, non-verbal, and contextual social information can be conveyed using CVEs than is possible with many other telecommunications systems" [SB02, p.237]. In IVEs, avatars representing interaction partners are experienced not as 2D images on a screen, but as life-size, 3D entities occupying a shared, surrounding mediated space. Figure 2.2 contrasts the experience of groupware systems such as videoconferencing, where the visual environment is on a a 2D screen separating the physical surroundings of each user, and IVEs, where it surrounds both users.

The VIRTUE [SS03] and SVTE [KS02] systems go some way in attempting to provide a 3D, surrounding experience without resorting to immersive technology. However, participants are constrained to a seating arrangement around a table, whereas CVEs provide a flexible interaction context.

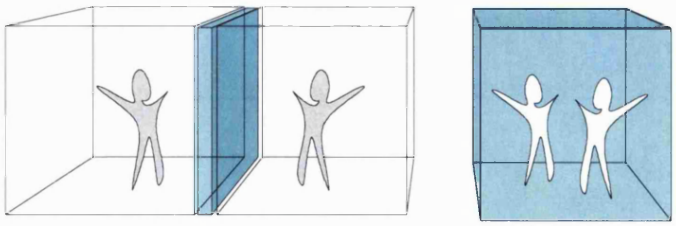


Figure 2.2: Using VMC, people remain in separate physical contexts and interact with each other via a video projection. Using IVEs, people interact in a shared, computer-generated 3D context where they are represented by digital proxies called 'avatars'.

2.1.3 Third dimension: fidelity

Fidelity concerns the degree to which objects and events in the mediated space are direct representations of the real world. Videoconferencing and media spaces portray participants' real appearance and actions, as well as views of their real environment. Conversely, CVEs portray

artificial, computer-generated scenes. In the context of group interaction, the degree of fidelity of a CVE hinges on its capacity to portray a convincing context and process for collaboration. This directly affects interaction with shared objects: "CVEs enable participants to work with shared access to objects located in the virtual environment, whilst media spaces endeavour to provide participants with the opportunity to work on 'real, physical' objects" [HFH+98, p.218]. Thus the advantage of CVEs is their ability to place objects in a 3D context, but their disadvantage is that the objects are not 'real.' Similarly, human embodiments in CVEs are synthetic and vary in the accuracy with which they mimic the real appearance and behaviours of the person they represent. This ability to couple anonymity with visual expressiveness has been cited as one of the hallmark attractions of online virtual communities [Lud96]. However, the ambiguous relationship between an avatar and the person represented also poses complex challenges in terms of creating expressive embodiments that contribute to the interaction taking place.

2.1.4 Summary

CVEs have several properties that make them suited to group interaction. They are:

- *spatial*, providing a shared 3D interaction context;
- *navigable*, allowing users to freely navigate the 3D space;
- *embodied*, representing users by digital proxies called 'avatars';
- *synchronous*, enabling people to interact with each other in real time;
- *multi-user*, supporting multiple, geographically dispersed users.

In summary, this section has discussed some of the potential advantages of CVEs as a communications medium. Their spatiality and immersiveness set them apart from other groupware systems in their ability to provide a surrounding, multisensory environment with consistent spatial properties. However, CVEs are by definition synthetic environments and therefore one of the challenges is to increase fidelity while preserving the advantages of spatiality and immersiveness.

2.2 State of the Art in CVEs

Numerous practical and collaborative uses are envisaged for CVEs, including virtual business meetings, scientific co-visualisation, virtual therapy and entertainment, teaching and collaborative design. However, CVEs have yet to come into mainstream use as a communications medium. CVEs range from text-only MUDs to fully immersive, multiuser 3D graphical environments. Though all are are based on a spatial model of interaction, MUDs provide verbal

descriptions of participants and places whereas graphical environments make them visually explicit in a consistent coordinate system where "relative positions and orientations of different objects can be measured" [BBRG96, p.78]. Table 2.2 categorises CVEs in order of increasing communicative bandwidth:

Table 2.2: Classification of CVEs

Interface	Type of CVE	Modality
Desktop	MUDs	Text-only interaction
	Graphical chats	Visual and text interaction
		Visual and audio interaction
Immersive	IVEs	Surrounding visual and audio interaction

This thesis focuses on the role of avatars, and as such is exclusively concerned with graphical environments. Schroeder points out that social interaction in CVEs has been studied in two distinct areas representing different research aims and approaches [Sch02b]. Research on virtual communities is typically characterised by longitudinal, naturalistic observations of large-scale virtual worlds. Conversely, laboratory research is characterised by controlled experiments on small group collaboration, often involving higher-end technical equipment. The research in this thesis belongs to this latter category. This section will briefly introduce graphical chat spaces to contextualise the problem of creating expressive avatars. Next, it will discuss related research on small group interaction in CVEs.

2.2.1 Online communities

A number of graphical chat environments have sprung up in recent years. Also known as *online environments* [SFD02], *on-line internet environments* [BDR00], *3D virtual worlds* [AV02] or *internet-based desktop 'social' VEs* [Sch02a], graphical chats are large-scale, networked VEs primarily designed for entertainment purposes. They share the general characteristics of CVEs in that they are spatially navigable, permitting multiple embodied users to interact with each other. They are also persistent over time, providing a space that participants can visit repeatedly [Jak02].

Avatars play an essential role in these environments. Whereas participants' presence in text-based MUDs is signalled by room listings or text interaction, Taylor argues that their presence in graphical chats is automatically articulated by avatars: "The avatar *as a body* is woven into the structure of life in these worlds. It is through embodied practice that selves and social life are grounded in multi-user spaces" [Tay02, p. 60]. Verbal communication in the *Palace* [Pal03] and *Activeworlds* [Act03] graphical chats is via text chat, whereas *Onlive!Traveler*

[Onl03] supports real-time audio interaction. In the Palace, avatars are 2D and users express themselves by alternating images from their repository of 'avatar' bitmaps. In contrast, *Onlive!Traveler* avatars are 3D and capable of facial animation and lip synch. Avatars are represented only as heads in order to reduce rendering costs. Conversely, users in *Activeworlds* are represented by full-body avatars but interact via text chat.

Research in online communities has focused primarily on longer-term communication issues concerning the role of persistent online identities and emerging social conventions in shaping virtual communities. Avatars have been discussed in relation to identity production [KKER98] and gender swapping [Lud96]. One of the hallmark attractions of graphical chats is the control afforded over users' visual appearance [Sul99], offering a safe environment in which to explore aspects of 'self'. However, less research has been conducted on the function of avatar behaviour in supporting the moment-to-moment unfolding of conversation.

One interesting behavioural observation in graphical chats concerns proxemics (spacing behaviour). Users engaged in conversation appear to mimic face-to-face rules for spatial behaviour, positioning their avatars in front of each other at an acceptable social distance. This finding is consistent across a number of studies using different CVE systems [BM02, SFD02, KLCH00, KKER98]. This emergent use of spatial framing offers compelling evidence that avatars serve as more than navigational aids.

Becker and Mark claim that socialising in graphical chats is indeed affected by nonverbal cues [BM02]. Avatars in *Activeworlds* are given automatic behaviours to make them appear lively. In addition, users can select a number of pre-programmed animations, though studies indicate that their use declines with habituation to the system [BM02, SFD02], suggesting a limited novelty value. This may be partly due to the cognitive load involved in keeping track of the text-based interaction. In addition, the animations themselves may not be appropriate to enriching communication. While it is important for avatars to appear lively, it is equally important that their behaviours convey the desired message.

A common critique of currently available graphical chat worlds such as *Activeworlds* is that avatars merely serve as placeholders, giving an indication of the individual's location and orientation within the shared space. Cassell and Vilhjálmsson argue that they fail to reflect the ongoing conversation: "While the user is creating the message for her interlocutor, her avatar stands motionless or repeats a selected animation sequence. This fails to reflect the natural relationship between the body and the conversation that is taking place, potentially giving misleading or even conflicting visual cues to other users. Some voice-based systems offer simple lip synching, which greatly enhances the experience, but actions such as gaze and gesture

have not been incorporated or are simply produced at random to create a sense of 'liveliness"' [VC99, p.45]. This signals a need to enhance the expressive capabilities of avatars by studying how nonverbal behaviours can contribute meaningfully to the ongoing communication process.

In the majority of cases, graphical chats are used for social and entertainment purposes. The Boeing company attempted unsuccessfully to use an online community as a shared interaction space for their distributed working groups. Fuchs, Poltrock and Wojcik suggest that its failure was due to employees' perceptions of avatars as a form of masquerade inappropriate to serious uses [FPW98], and suggest that avatars need to convey richer and more reliable information before they can be used for business purposes.

2.2.2 Research on small group interaction in CVEs

Schroeder has said that what makes CVEs different from single-user VEs is that users are doing things together in the environment, or are at least aware of each other [Sch02a]. The problem of illustrating awareness through avatars in a spatial model of interaction is challenging, and has been discussed in some depth by Benford and colleagues [BBF+94]. They propose mechanisms for handling awareness through the use of aura (the space bounding an avatar's presence), focus (the space a user is aware of) and nimbus (the space in which others can become aware of the user's avatar). They stress the importance of the avatar in signalling each user's focus of attention, activity, and potential availability for interaction.

A later paper by Bowers, Pycock and O'Brien presents an empirical evaluation of small-group interaction in a CVE, using conversation analysis to investigate the effect of user embodiment on turntaking [BPO96]. The premise is that avatar position and orientation should theoretically support mutual awareness. In practise, users did position their basic 'blocky' avatars to face the people they were speaking with, using whole-body rotations to indicate 'glances' at a speaker; this evidence of spatial interaction is consistent with the research findings from graphical chat spaces mentioned in Section 2.2.1, and suggests again that avatars "have interactional significance and are not merely navigation devices" [BPO96, p.64].

Hindmarsh et al. further investigated the role of avatars in a study on object-focused collaboration in a shared design task [HFH+98]. Though visually simple, the humanoid avatars raised expectations for humanlike functionality, leading participants to misjudge their interaction partners' field of view (FOV). Hindmarsh argues this leads to a'fragmentation' of the shared space, with participants unable to clearly establish each other's attentional focus. The problem is compounded by the limited horizontal FOV afforded by the desktop interface, and the limited functionality of the avatars used. Nevertheless, the interesting finding is that even basic embodiments raise expectations for human functionality.

Slater, Tromp, Steed and colleagues ran three experiments on small group interaction in a CVE [SSUS00, TBS+98, SSS+99]. Trios of strangers met in a VE to perform a puzzle-solving task involving the reconstruction of idiomatic expressions from individual words distributed on the walls. They were represented by avatars labelled Red, Green and Blue. The avatars had limited and identical functionality, showing each participant's position and orientation but no additional facial or bodily movement. Participants could see the two avatars representing their collaborators, but did not know how they themselves were represented. The experiments addressed the following questions:

- Whether computational advantage (having higher-end equipment) confers social power in the context of small group interaction;
- Whether there is an association between the sense of presence (being in the virtual place) and the sense of copresence (being with others);
- Whether VEs can elicit a sense of social discomfort.

Some key differences between the three experiments are summarised in Table 2.3:

Table 2.3: Comparison of three small group interaction experiments by Slater, Tromp, Steed and colleagues

	Experiment 1 [SSUS00]	**Experiment 2** [TBS+98]	**Experiment 3** [SSS+99]
Number of groups	10	4	20
Location of participants	London	Distributed: London, Nottingham, Sweden	Distributed: London, Nottingham, Greece
Number of immersed participants	1: Red (Green and Blue used desktop)	None	1: Blue (Red and Green used desktop)
Avatar appearance	Same for all three participants	Green was more photo-realistic. Red and Blue had radio waves emanating from their heads while speaking	Same for all three participants
Findings: relationship between presence and co-presence	Significant positive correlation	No significant relationship	Significant positive correlation
Findings: relationship between computational advantage and social power	Immersed participant significantly more likely to be perceived as leader	Not applicable	No straightforward relationship

Regarding the relationship between computational advantage and social power, the overall results are ambiguous. In the first experiment, participants began a puzzle task in the VE,

filled out a questionnaire, and continued the task in the physical room the virtual space was modeled on. One participant (Red) used an HMD, whereas Green and Blue used a desktop interface. The analysis of questionnaire data suggested the immersed participant was significantly more likely to be perceived as the leader; however, this effect disappeared when the task was continued in the real environment. In the second experiment, participants were distributed. This time, all used a desktop, in order to investigate whether the earlier finding regarding perceived leadership was attributable to Red's immersive interface in the first experiment. The fact that no significant leadership pattern emerged lent support to this hypothesis. However, the third (distributed) experiment found no straightforward leadership pattern, leaving this question unresolved. Interestingly, immersed participants did not report a significantly higher level of experienced presence in either the first or third experiment.

Both experiments 1 and 3 found a significant positive correlation between presence and copresence. This was not the case in experiment 2, although this was based on a significantly smaller sample size (only four groups as opposed to 10 groups in experiment 1). Regarding social discomfort, a potentially awkward social situation was engineered in experiment 1 by instructing Green to always remain in Red's line of vision. No significant quantitative effect was found resulting from Green's monitoring of Red, although qualitative findings from the debriefing sessions revealed that the social dynamics were affected in a variety of ways, with some participants feeling a sense of hostility, discomfort or exclusion.

One qualitative finding from experiment 2 is particularly relevant to considerations of avatar fidelity. Here, Red and Blue were represented by simple 'blocky' avatars with little visual detail, while Green was represented by a more photorealistic one (see Figure 2.3). The debriefings illustrate that differences in appearance affected responses. In one case, the higher-realism Green avatar was thought to represent a figure of authority. In another it was seen as "scary, like a zombie." In yet another case, the two lower-realism avatars were seen as cartoonish and were assumed to be 'robots' rather than real people. Tromp et al. [TBS+98] conclude that higher realism in an avatar's appearance may lead to heightened expectations for behavioural realism. This crystallises the need to further explore the relationship between the appearance of an avatar and its behaviour.

The lack of visual feedback from the avatars was seen as a barrier to interaction in all three experiments: "Lack of eye-contact, body language, and even the ability to point at a reference object were important drawbacks" [SS02]. In particular, the lack of eye gaze was problematic because participants were unable to monitor each other's directed attention in the shared task. Nevertheless, the qualitative findings from these studies illustrate that despite their simplistic

Figure 2.3: Avatars used in the experiments by Slater, Tromp, Steed and colleagues

appearance and minimal behaviours, the avatars were able to elicit social responses. Participants were respectful of the avatars, for example expressing distress at accidentally walking through each other.

Schroeder, Wideström and colleagues extended Slater et al.'s work in a series of studies examining the impact of different display technologies on people's interaction experiences on a collaborative puzzle-solving task [SSA+01, WAS+00]. A highly interactive object-focused task was used, involving the reconstruction of a 3D object similar to a Rubik's cube. Pairs of participants were given a limited time to reconstruct the cube under the conditions summarised in Table 2.4:

Table 2.4: Interfaces used in collaborative puzzle cube studies by Schroeder, Wideström and colleagues [SSA+01, WAS+00]

STUDY	PARTICIPANT 1	PARTICIPANT 2
Study 1 (Symmetrical)	Face-to-face	Face-to-face
Study 2 (Asymmetrical)	Immersive: 5-sided Cave	Non-immersive: Desktop
Study 3 (Quasi-Symmetrical)	Immersive: 5-sided Cave	Immersive: 4-sided Cave

One of the purposes of the studies was to investigate the impact of computational advantage on a range of responses including collaboration, presence and copresence. Both symmetrical and asymmetrical collaboration conditions were investigated. Since these studies involve pairs rather than groups of three, Schroeder et al. defined leadership in terms of the degree of each participant's perceived contribution to the task. In the second study, participants in each pair consistently agreed that the immersed person contributed more to the task than the non-immersed person, despite the fact that they were unaware of any differences in their setups. In both the other studies, where the setup was more symmetrical, participants felt they had

contributed equally, further supporting Slater et al.'s earlier findings [SSUS00] regarding the correlation of computational advantage with leadership. Interestingly, participants in all studies felt that their verbal contributions were equivalent; only the object manipulation activity was affected by the asymmetry in the second study.

As expected, reported presence was significantly higher for immersed participants than for those using the desktop. The interesting finding concerns copresence (defined by Schroeder as the sense of being together in the computer-generated environment). Reported copresence was significantly higher in symmetrical situations where two immersed participants worked together than in asymmetrical situations where an immersed participant was paired with a non-immersed partner. This suggests that participants' sense of copresence is strongly related to the sense of presence reported by their partner.

Participants in these studies were represented by a visually simplistic male avatar. Participants in the immersive setup saw their partner as life-size, but only saw the arm of their own avatar so they remained unaware that they were visually identical. The problem-solving tasks used in these studies were highly object-focused rather than personal, and this partially compensates for the relative paucity of the avatars. There is evidence [WO97] that people need less visual feedback from each other in tasks focusing on an object. However, in the more interpersonal interactive tasks envisaged by Schroeder et al.'s discussion of future work, nonverbal behaviours would likely play a more significant role. For this reason, the experiments discussed in Chapters 4 and 5 of this thesis used a negotiation task. The research presented in this thesis focuses on the avatar's role in social interaction, and it was necessary to create a context in which participants would ordinarily place greater reliance on nonverbal feedback.

2.3 Research Problem: Increasing Avatar Fidelity

One significant barrier to interaction in current CVEs is in the paucity of avatar expression compared with live video of real people. One of the challenges in developing CVEs as a communications medium is therefore the creation of expressive avatars. This section will begin by defining and classifying avatars. It will then discuss some communication requirements and the challenges these entail. In particular, there are technical restrictions on the amount of visual detail that can be conveyed and on the ability to drive appropriate behaviours in real time. Increasing avatar expressiveness therefore entails a potential tradeoff between photorealism and behavioural realism. This section discusses related research studies on the impact of different aspects of appearance and behaviour on people's social responses to avatars and agents. It concludes with a discussion of the approach taken in the research, namely to investigate the impact

of minimal fidelity on people's social responses to virtual humans.

2.3.1 Classifying virtual humans: avatars and agents

Virtual humans are visible, computer-generated humanoid characters used for a wide range of applications. They can function as interface agents, news readers, game characters, digital extras populating film sets and archaeological reconstructions, surrogates for medical training, and as personalised dummies used to try on clothes in virtual shopping applications.

By convention, virtual humans are classified in terms of agency, meaning whether the intelligence represented is human or artificial [SB02, Bla02]. Where avatars represent real humans engaged in interaction, agents are driven purely by a computer program and can vary widely in sophistication. Some agents have simple, pre-scripted behaviours whereas others such as MIT's virtual estate agent, Rea, are designed to sustain verbal and gestural interaction with human interlocutors [CBB+99]. Research into embodied conversational agents is driven by fundamentally different concerns and is therefore beyond the scope of this thesis, which focuses on human-to-human communication.

In the strictest objective sense, agency is binary because the virtual human either represents a human or it does not. Nevertheless, avatars vary in the degree to which their behaviours represent the real actions or intentions of the person represented. Blascovich argues that agency is a continuum ranging from fully artificial at the low end, to fully human at the high end, with the term 'avatar' being reserved for the upper extreme of the continuum. Blascovich stresses that agency is subjective, in that it is "the extent to which individuals perceive virtual others as representations of real persons" [Bla02, p.130].

This thesis takes into account both the objective and subjective views. Firstly the objective, because virtual humans used in the three experiments presented are strictly classified according to whether or not they represent human participants. Visually identical virtual humans are used in two of the experiments presented in Chapters 4 and 5; when driven by pre-scripted behaviours they are referred to as agents, and when representing humans they are referred to as avatars. Their level of behavioural complexity is not taken into consideration when making this distinction. Secondly the subjective, because independently of their objective status, the degree of sentience attributed to virtual humans can vary depending on their behaviour in the course of the interaction. The question of how attributed sentience varies with the visual behaviour of virtual humans is addressed in the experiment presented in Chapter 4. The notion of agency and its relationship with virtual human behaviour is brought up again in the discussion of Blascovich and colleagues' model of social influence [BLB+01] in Section 2.5.6.4.

The challenge with human-human communication is to drive avatar behaviours that en-

rich, rather than hinder, communication between remote participants. The following subsection addresses some communication requirements, and is followed by a discussion of some key technical constraints shaping the development of expressive avatars.

2.3.2 Goals for expressive avatars

Benford et al. have laid out some requirements for avatars in CVEs covering aspects of appearance, behaviour and relationship to the real body of the person represented [BBF+95]. For clarity, the key visual and behavioural requirements are classified in Table 2.5:

Table 2.5: Summary of avatar requirements, after Benford et al. [BBF+95]

	Requirement to convey	Comment
Appearance	Presence	Must signal person's presence in the VE in an "automatic and continuous" way
	Location	Must signal the person's position and orientation in the VE
	Identity	Avatar's appearance must provide recognisability over time as well as the ability to distinguish between avatars
Behaviour	Availability	Must convey person's availability for interaction, and how busy or interruptible they are
	Activity	Must portray person's current activity and focus of attention
	Expression	Should convey expressiveness through gesture and facial expression

Benford et al. argue that that these requirements often conflict with each other, and prioritisation hinges on interaction context and technical resources [BBF+95]. It is arguable, however, that requirements cannot be easily separated because the avatar's appearance and behavioural requirements often intertwine. For instance, simple T-shaped 'blocky' avatars are sufficient to signal presence and location, and their colour can identify them as distinct from other avatars. However, for other functions the avatar is likely to require more visual detail, such as eyes to convey attention and arms to convey simple interaction with objects such as grasping (as implemented by Hindmarsh et al. [HFH+98] in their study on collaborative object manipulation). These functions, though challenging, are relatively simple compared to the difficult problem of conveying convincing behaviour.

In face-to-face interaction people rely heavily on nonverbal cues such as eye gaze, facial expression, posture, gesture and interpersonal distance to supplement the verbal content of conversation. Indeed some argue that nonverbal signals not only constitute a separate channel of communication, but that they often override verbal content [For85]; in other words 'how' something is said can be more important than 'what' is said. Thomas and Johnston emphasise that the need to maintain consistency between dialogue and nonverbal expression is equally impor-

tant in cartoon animation: "Do not let the expression conflict with the dialogue. Nothing can be more distracting than this" [TJ81, p.441]. This points to a need to align the visual behaviours of avatars to the ongoing interaction. A brief discussion of salient nonverbal behaviours follows.

2.3.3 The importance of nonverbal communication in face-to-face interaction: a brief review

Nonverbal behaviours serve at least two central functions in face-to-face interaction: conversation management and the communication of emotion. Conversation management concerns the use of paralinguistic cues to ensure the smooth flow of conversation. Movements such as eyebrow raises, head nods and posture shifts give structure and rhythm to the conversation and are essential to maintaining a sense of mutual understanding. The communication of emotion is itself integral to the regulation of communication and interaction [Pic97, Gol96]. In the words of Picard, "emotions not only contribute to a richer quality of interaction, but they directly impact a person's ability to interact in an intelligent way" [Pic97, p.2]. Emotion is crucial in the communication of understanding, and speakers continually monitor listeners' body language and facial expression for confirmation that they are being understood.

2.3.3.1 Facial expression

Within nonverbal communications research the greatest amount of attention has been devoted to facial expression, possibly because there is considerable consensus that the emotional signals are most efficiently conveyed through the face [Iza97]. Researchers in the Darwinian tradition believe that emotion is the result of evolutionary processes [Cor96] and that there are therefore several aspects of emotional communication which are universal across cultures. Ekman [EFE72], Izard [Iza71] and others have agreed on a set 6 'primary' emotions that can be decoded well above chance from facial expression alone: happiness, sadness, surprise, fear, anger, disgust and contempt.

Although these results have been challenged on methodological grounds, primarily because they used static photographs and forced-choice questionnaires, recent findings suggest that the same six basic emotions can be reliably decoded in moving video as well as computer-generated characters [ES00]. In addition to the notion of pan-cultural expression, Ekman discusses the importance of socially-learned patterns of behaviour called 'display rules'. These operate in four distinct ways: de-intensification, over-intensification, neutral or affectless appearance, and masking one emotion with another.

True emotions are often masked to comply with social expectations. In instances of 'nonverbal leakage', a true emotion will accidentally be conveyed despite efforts to mask it. Gener-

ally, awareness and control of facial expression is keener than for bodily behaviours. Goleman argues that people are not usually aware of their nonverbal behaviour [Gol96]. Ekman and Friesen [EF69] suggest that the activation of display rules becomes a force of habit and typically occurs below the level of awareness. In theory at least, effective control of an avatar's behaviour would allow users to transmit only their intended meaning, with a complete elimination of nonverbal leakage; the challenge in developing expressive avatars concerns the transmission of intended cues without requiring users to attend to driving their avatar's nonverbal behaviours.

2.3.3.2 Eye gaze

Eye gaze is a richly informative behaviour in face-to-face interaction. It serves at least five distinct communicative functions in conversation: [AC76, Ken67]:

1. regulating conversation flow;
2. providing feedback;
3. communicating emotional information;
4. communicating the nature of interpersonal relationships;
5. avoiding distraction by restricting visual input.

The perception of eye gaze depends on a combination of head and eye orientation [AC76, AC76, GP63]. Bruce and Young [BY98] also point out that the whites of the eyes stand out in humans, making the direction of gaze easy to determine. In most animals the white area of the eye is minimised to avoid unnecessary conflict, because direct eye contact is considered a threat signal. A long gaze (or stare) is considered a signal of dominance or aggression even among humans, but the appropriateness of gaze duration is dictated by social conventions and contexts [Arg88, BY98]. Longer gaze duration can also be a sign of intimacy, for instance between romantically involved people [BY98].

Disney animators discovered that in order to make animation convincing it is necessary to consider the overall pattern of facial and bodily movement rather than individual features [TJ81]. However, if considering the face, the eyes were found to be the single most expressive feature. Eye shape expresses emotion, and the direction of gaze indicates to what or whom that emotion is directed.

Gaze is also an important precursor to interaction, regulating the beginning and ending of social encounters through the "making and breaking of mutual gaze" [Arg88, p.161]. During conversation, it continues to act as a two-way channel, allowing speakers to monitor others for

attention and understanding, and listeners to signal interest and attention. It therefore plays a central role in the portrayal of directed attention. Gaze also plays an important role in the regulation of turntaking. Typically a speaker will make longer eye contact towards the end of his turn, often selecting the next speaker by ensuring that a mutual gaze is established with that person [Ken67].

2.3.3.3 Body movements: gesture, posture and proxemics

Gesture, posture and proxemics have received less research attention than either facial expression or gaze. Ekman and Friesen [EF69] divide gestures into different categories according to their communicative functions:

1. *emblems* are used consciously and intentionally, and usually have a culturally codified meaning that can be substituted with a word or phrase, such as 'thumbs up';
2. *illustrators* such as baton signals are directly tied to speech on a moment-to-moment basis, and are used to emphasise the rhythm of spoken dialogue;
3. *regulators* are used to mark the flow of the conversation as a whole, for example in indicating the next speaker with a hand gesture.

There is evidence [EF69, Wal98] that the body can communicate information about emotion on several levels. Ekman and Friesen [EF69] suggest that while the face communicates information about the nature of an emotion, body movements ('acts') convey additional information about the intensity of an emotion. Further, still positions (postures) can communicate information about intensity and sometimes gross affective state along a pleasant/unpleasant dimension.

Posture changes at a slow rate, and is therefore more relevant to longer-term aspects of conversation rather than to micro-momentary feedback. Argyle [Arg88] ties posture to the expression of mood and personality. Bull [Bul83] identified a link between postures and certain emotions. Interest is associated with a forward lean and drawing legs back, whereas boredom is associated with a backward lean, lowering of the head or leaning the head on one's hand, outstretched legs, and turning the head away.

Proxemics, or spacing behaviour, concerns both interpersonal distance and orientation. Argyle and Dean's intimacy equilibrium theory suggests an inversely proportional relationship between gaze and interpersonal distance [AD65]. If distance is decreased, as in the confinement of a lift, people will lower or avert their gaze to maintain the same overall level of intimacy. Also linked to interpersonal distance is the notion of polite inattention, a distance of 12-15

feet recognised as the range beyond which it is considered acceptable not to acknowledge another person's presence [SS72]. Hall presented a theoretical model for interpersonal distance, consisting of four concentric zones indicating decreasing levels of intimacy [Hal66]. Different behaviours are considered appropriate for each zone. This model was used in the experiment on responsiveness in Chapter 4 to delineate different response zones for the agents; the agents' behaviours were modified as a function of the participant's distance from them.

In addition to interpersonal distance, orientation also plays an important role in signalling availability. Scheflen and Scheflen suggest that the angle at which people position themselves in relation to each other forms either closed or open frames [SS72]. Frames communicate whether a group is engaged in private discussion or welcomes new members. Orientation, like gaze, helps to regulate the beginning and ending of conversation by facing towards or away from conversation partners.

In summary, nonverbal behaviours play a central function in face-to-face conversation. Avatars' ability to convey such nonverbal cues is likely to affect how they are perceived as well as their contribution to social interaction. In works of cyberfiction such as Neal Stephenson's *Snow Crash* [Ste92], avatars are both highly photorealistic and expressive. They perform seamlessly in real time, and are so reliable in conveying intended behaviour that businessmen happily substitute face-to-face meetings with interactions in the 'Metaverse.' The reality of CVEs today is strikingly different. The following section will contextualise avatar appearance and behaviour in current CVEs in terms of the technical constraints on fidelity.

2.4 Constraints on Avatar Fidelity

There are key technical constraints affecting the degree of avatar fidelity possible in current CVEs. In this thesis fidelity is taken to encompass both static properties of avatar appearance (visual fidelity), and dynamic properties of animation (behavioural fidelity). The first consideration, with regard to visual fidelity, is the tension between 'realism and real time' [SSC01]. Slater et al. individuate three aspects of realism in VEs: geometric realism, illumination realism and behavioural realism [SSC01]. While all these are desirable in the creation of convincing VEs, they come at the expense of real-time performance. In terms of an avatar's appearance, increased photorealism introduces computational complexity, resulting in significant and unwanted delays to real-time communication. The second consideration, regarding behavioural fidelity, is the tension between control and cognitive load. Mapping a person's communicative intentions to their avatar's behaviour presents considerable technical challenges. Full manual control introduces unacceptable cognitive load; on the other hand, reducing cognitive load

through tracking or alternative approaches results in a loss of full control over the avatar. Figure 2.4 summarises these tensions. Each will be discussed in turn with reference to visual and behavioural fidelity, respectively.

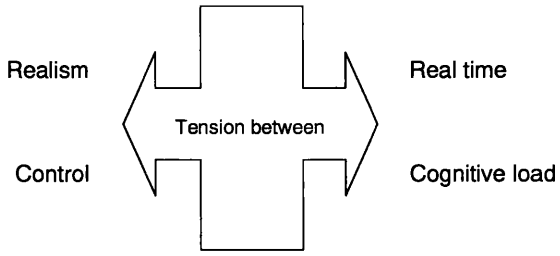


Figure 2.4: Technical constraints affecting avatar fidelity

2.4.1 Constraints on visual fidelity: the tension between realism and real time

In the novel *Snow Crash* [Ste92], photorealistic and personalised avatars are a sign of status in the virtual world. Schroeder argues that avatar embodiment affects how people relate to each other in CVEs, and that "avatar appearance will influence interaction in all shared VEs, and there is still much research to be done on pinning down this influence" [Sch02b, p.xi]. Findings reported by Nilsson et al. suggest that avatar appearance may not be as important for long-term collaborations, particularly where participants already know each other [NHSA02, p.120-1]. Nonetheless, in the context of one-off interactions of interest in this thesis, avatar appearance is likely to have some significance.

In terms of appearance, Schroeder points out that "it is not only the shape of virtual bodies that matters in the experience of virtual worlds, but also the level of detail with which they are represented" [Sch96, p.64]. Fidelity concerns not only morphology and photorealism, but also the degree to which the avatar resembles the person represented (referred to by Benford et al. as 'truthfulness' [BBF+95]). Figure 2.5 illustrates three dimensions of visual fidelity.

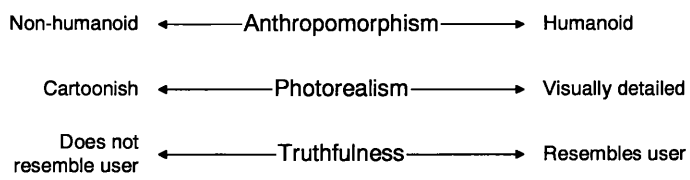


Figure 2.5: The dimensions of visual fidelity include anthropomorphism, photorealism and truthfulness.

In terms of photorealism, avatars can range from simple 'blockies' to highly realistic forms. Avatar morphology in graphical chats ranges from humanoid to anthropomorphised

animals to abstract shapes [Sul99]; the research presented in this thesis is concerned exclusively with 3D, dynamic humanoid avatars. Within this humanoid category, avatars can also vary in terms of their fidelity to the user's real-life physical appearance.

Typically, avatars used for communication purposes are relatively cartoonish. Cheng, Farnham and Stone suggest that users may prefer to be represented by humanoid avatars that are neither too cartoonish nor too photorealistic [CFS02, p.99]. The reason why highly photorealistic avatars are not used, however, is primarily due to technical constraints on local rendering and network bandwidth. Morningstar and Farmer cite the latter as a particular concern in the design of graphical chats, emphasising that "communication bandwidth is a scarce resource" [MF90]. Similarly Hindmarsh et al. advocate using recognisable but simplistic humanoid avatars for performance reasons: "We adopted this approach because we felt that it is the most obvious choice and indeed, is one that has been widely adopted by CVE designers" [HFH+98, p.218]. Their avatars had a head, torso and arms, and were capable of simple behaviours including looking, pointing and grasping objects.

2.4.2 Constraints on behavioural fidelity: the tension between control and cognitive load

Where visual fidelity concerns the static properties of an avatar's appearance, behavioural fidelity concerns its dynamic properties of animation. The primary focus of this thesis is on the *perception* of avatar behaviours rather than how they are *driven*. Nevertheless, the difficult problem of driving appropriate behaviours is of interest because it directly shapes the research problem. Research on nonverbal behaviour in face-to-face communication [Arg88] can offer valuable leads on how to improve avatar expressiveness without resorting to full tracking. This subsection discuss the problem of how to implement nonverbal behaviours in humanoid avatars.

As mentioned in Section 2.2.1, avatars in existing graphical chats have been widely critiqued for their insufficient and sometimes misleading behaviours. Durlach and Slater suggest that CVEs create a web of relationships connecting people to each other, and individually to their own avatars [DS00]. The way individuals relate to their own avatar is likely to hinge on how reliably it represents them. In enriching the communicative potential of avatars it is essential not to misrepresent the actions or intentions of users.

Avatar behaviours can be driven in a variety of ways, some of which are summarised in Table 2.6. As mentioned earlier in this section, full control over their behaviour usually comes at a cost in terms of cognitive load. On the opposite extreme, full automation can result in a loss of control over the avatar's actions.

Manual driving through menu selection, mouse movement or pen gesture affords control

Table 2.6: Approaches to driving avatar behaviours

	Full Control	Full automation
General approach	Manual input changes avatar state	Person's real movement changes avatar state
Examples	• Menu selection • Mouse input • Keyboard gesture • Pen gesture • Hand gesture	• Video textured onto avatar • Automatic face processing • Real-time whole body tracking, including face and eyes

over the avatar's actions. Systems using pen gesture [BC02] and hand gesture [LGPW98] have also been proposed. These approaches require continuous attendance to the avatar's state. For example, the solution adopted in Slater et al.'s acting rehearsal experiment used a combination of input techniques [SHS+00]. Actors positioned their avatars using keyboard arrows, and could raise their arms using a visual slider. They manipulated a selection of expressions with varying intensity by drawing simple, intuitive abstractions of eyebrow patterns, smiles or downturned mouths on a blank 'face' canvas. While the expressions were straightforward, the requirement to attend to the interface introduced cognitive load that at first detracted from their rehearsal task.

The problem of cognitive load is compounded in those graphical chats involving text interaction rather than live audio. For instance, users in *Activeworlds* tend to focus their attention on typing and following dialogue rather than on the graphics [MC98]. Cheng, Farnham and Stone cite cognitive load, along with technical restrictions, as major detractors to the development of graphical chats: "Despite advances in technology over the past six years, multi-user 3D environments still have difficulty achieving critical mass, particularly in scenarios for practical applications (as opposed to those for socialising). We believe that for many communities, the demands on the users' attention and the networking and machine requirements will continue to be a barrier to building critical mass" [CFS02, p.110].

Several alternative approaches have been proposed in response to the problem of enriching avatar communication while reducing cognitive load. Cuddihy and Walters suggest a solution involving high-level control through a dynamic interface that clarifies what actions are available to users at any given time [CW00]. This would make it possible to direct a 'waving' action at an approaching avatar rather than manually orienting the avatar and then raising its arm, as was

the case in [SHS+00]. Tromp and Snowdon suggest the automation of behaviours to enhance group interaction, for instance locking gaze to the speaking avatar to denote attention [TS97]. The drawback is that automation may result in misleading behaviours. In the second experiment presented in Chapter 5 it was decided not to automate gaze locking precisely because it was believed that mutual gaze should arise from participants actually facing each other's avatars.

A similar high-level approach is taken by Vilhjálmsson and Cassell in the BodyChat system [VC98]. Here, users choose whether to be available for conversation, and their avatars automate appropriate cues such as smiles, eyebrow raises and glances to indicate a willingness to approach or depart. In a within-subjects evaluation, they compared subjective responses to three systems: this automated approach, fully manual driving, and a system that combined the ability to drive and automate behaviours. Their analysis of questionnaire responses suggests that the automated condition received significantly higher ratings in terms of perceived naturalness and expressiveness of the system, and ability to remember details about interaction partners. Surprisingly, participants also reported a greater sense of 'control' using the automated system, which the authors attribute to a sense of greater control over the conversation as a result of being freed from detailed attendance to their own avatar.

These approaches have involved reducing cognitive load from manual input interfaces. A radically different approach involves mapping the person's real-life expression to the avatar's. Durlach and Slater indicate two possible approaches: the use of "direct, pass-through video of the participants" [DS00, p.216], or using tracking data to manipulate the avatar's 3D mesh. Tracking has both advantages and disadvantages. It involves a reduction in cognitive load, however head and wand trackers can be invasive for participants to use as well as expensive in terms of equipment and rendering. Tracking theoretically allows for the transmission spontaneous expressions, which Benford et al. cite as a particularly challenging problem in CVEs [BBF+94]. The degree to which involuntary expression is desirable is debatable in a medium that is valued for the control it gives users over the appearance and actions they convey to others.

If the goal is to replicate each person's real movement, tracking can seem an attractive solution. Systems such as Eyematic [Eye03] have shown compellingly that it is possible to track eye movement and drive an avatar in real time using a simple desktop camera. As Vilhjálmsson and Cassell have argued, however, the use of gaze tracking in non-immersive systems becomes problematic in group interaction because the user and avatar occupy radically different spaces [VC98]. Similarly to PIP videoconferencing, "the gaze pattern and orientation information gathered from a user looking at a monitor does not map appropriately onto an avatar standing in a group of other avatars" [VC99, p.4]. Immersive systems reduce the problem of spatial

mapping, but where users' faces are obscured by HMDs or stereoscopic goggles it can be challenging to provide a satisfactory solution.

Overall, there are significant challenges in driving appropriate behaviours for avatars. In addition to technical challenges, there remain open questions about the appropriateness of tracking or automating behaviours in the quest to reduce cognitive load without sacrificing users' control over avatar actions. Though not the focus of this thesis, the challenge of driving avatar behaviours shaped the approach taken in the two experiments on eye gaze presented in Chapter 5. In both cases, there was no suitable tracking solution for driving an avatar's eye movement in real time from the person's real eye movement. Given the impossibility of manual driving, the approach taken involved inferring eye movement from the audio stream based on rules for conversational turntaking in face-to-face interaction. Here, the goal was to explore a potentially inexpensive solution to improving avatar fidelity without resorting to manual control or full tracking.

2.4.3 Prioritising aspects of avatar fidelity

The previous subsections underlined the technical constraints on avatar fidelity in current CVEs. The tension between realism and real time limits visual fidelity, and the tension between control and cognitive load poses difficulties for driving high-fidelity behaviours. Given these constraints, this subsection will contextualise the approach taken in this thesis by discussing the current need for tradeoffs in developing expressive avatars. It will also present related research suggesting that avatars and agents can elicit social responses even given minimal fidelity.

Fraser et al. have stated that "virtual environments–models, avatars, interfaces and so on–are often designed with realism in mind" [FGV+00, p.30]. The 'Marilyn' avatar from Miralab is an example of a highly photorealistic avatar with realistic musculoskeletal functionality and realistic skin [Mar03]. The underlying assumption appears to be that more realistic environments and avatars should result in qualitatively better experiences in CVEs. Schroeder argues that this assumption needs empirical validation and lists a series of testable hypotheses, including one that directly concerns avatar fidelity: "In relation to the realism of the representation of the other person, the more realistic the appearance of the other person, the higher the copresence (or 'social presence')" [Sch02a, p.283][3]. This question will be directly addressed in the second experiment described in Chapter 5, where the impact of both appearance and behaviour are investigated in terms of their impact on social responses.

The need for literal portrayals in VEs is a matter of debate. As Zeltzer argues, given current

[3]Note that the use of the terms copresence and social presence will be addressed in greater detail in Sections 2.5.4 and 2.5.6.

technical limitations, the priority is to develop selective fidelity based on contextual needs: "It is not possible to simulate the physical world in all its complexity and detail, so for a given task we need to identify carefully the sensory cues that must be provided for a human to accomplish the task, and match as closely as possible the human perceptual and motor performance required for the task" [Zel92, p.128]. He goes on to state that further research is needed to understand how to measure selective fidelity. Similarly, Fraser et al. propose a shift in priorities away from literalism and realism, particularly given the crudeness of current interfaces for conveying human movement [FGV+00].

Benford et al. argue that improving avatar expressiveness necessarily involves compromises: "Designing a virtual body will therefore involve identifying the most important issues and trading them off against the available computing resources" [BBF+94, p.658]. They expand on this in a later paper, stating that computing and communication resources will always be limited, and highlighting the need to streamline avatars: "We suspect that approaches which attempt to reproduce the human physical form in as full detail as possible may in fact be wasteful and that more abstract approaches which reflect the above issues in simple ways may be more appropriate" [BBF+95, p.245]. Benford and colleagues advocate incremental context-driven improvements to fidelity rather than an absolutist drive towards photorealism.

Several authors share the alternative assumption that for communication purposes, behavioural fidelity is the higher priority. For instance, Sallnäs argues that realistic appearance is secondary to the support of body positioning, pointing and object manipulation necessary in collaborative tasks [Sal02]. Blascovich reasons that "because we typically build digital IVEs, including interpersonal ones, using visual media, we tend to think of realism in terms of photographic realism. Although important, photographic realism does not equate with behavioural realism and is, in fact, less important" [Bla02, p.131]. In a separate paper with Swinth, he adds that "more important than photorealism, and perhaps even anthropomorphism, is an avatar's behavioural realism. Behavioural realism refers to the extent to which avatars and other objects in an virtual environment behave like their counterparts in the physical world" [SB02, p.329].

The assumption that visual fidelity is secondary to behavioural fidelity is partly supported by lessons from animation. Disney animators translated films of actors' body language and facial expression into simple line drawings and discovered it was possible to achieve effective emotional portrayals in visually simplistic characters, provided the movement was convincing [TJ81]. More recently, Katsikitis and Innes' [KI97] study on line drawings of a smile illustrated that even a cartoonish representation of an expression can be decoded accurately down to its five phases of development.

Recent studies on the transmission of nonverbal cues in mediated communication add further support to the argument favouring behavioural fidelity. Ehrlich and Schiano point out that the same bandwidth restrictions constraining CVEs also apply to VMC [ES00]. They suggest that the standard approach of preserving spatial and colour resolution at the expense of temporal degradation is counterproductive. Their experimental findings indicate that preserving motion information is critical to the recognition of facial expression and may compensate for significant losses in image resolution. Table 2.7 summarises their findings regarding the impact of degradation on facial and affect recognition.

Table 2.7: The role of visual and temporal resolution in transmission of facial affect (after Ehrlich and Schiano [ES00])

	Facial recognition (appearance)	**Affect recognition (behaviour)**
Associated with	Image quality (spatial and colour resolution)	Visual dynamics (temporal resolution)
Effect of degradation	Robust across visual degradation	Sensitive to temporal degradation

Considering that the transmission of nonverbal cues can be severely affected by temporal delays and inconsistencies, they suggest that "if a bandwidth tradeoff is required, one should consider preserving high-fidelity motion information at the expense of image realism, not the other way around" [ES00, p.252]. In a separate study on facial affect recognition, Schiano, Ehrlich and Krisnawan compared a low-fidelity robot enacting the six 'basic' emotions with video of human actors enacting the same emotions [SEKS00]. Though scores for the robot were lower, the expressions were decoded in a pattern that closely followed the human faces. This further supports the argument prioritising behaviour over accurate appearance in the transmission of nonverbal cues. Bente and Kramer [BK02] describe a related study on person perception, this time comparing silent video clips of dyadic interactions between human actors with equivalent clips of identically animated agents. Their findings indicate a remarkable correspondence in responses to the video and agent conditions, despite the lower-fidelity appearance of the agents.

In summary, technical limitations have forced the need to set priorities in avatar design. In the words of Heeter, "Faced with technological limitations which prevent being able to simultaneously simulate all aspects of human perception, the alchemy of presence in VR is in part a science of tradeoffs. Which elements are most critical to the experience of presence? When forced to choose between responsiveness to motion and resolution of images, developers are choosing responsiveness as the more important factor, based on their own experiences and ob-

servations of others" [Hee92, p.264]. These findings from different media experiences partially support the notion that behavioural fidelity may be more pressing than visual fidelity. This is supported by Tromp et al.'s suggestion that higher-realism avatars in their experiment appeared to raise higher expectations for humanlike behaviours, suggesting that appearance should remain minimal until behaviour is sufficiently sophisticated to satisfy expectations [TBS+98].

2.4.4 Exploring the impact of minimal fidelity

The argument for exploring the lower boundaries of fidelity is not born exclusively out of technical necessity. Reeves and Nass document a series of studies suggesting that people respond to media as social actors, and tend to anthropomorphise even the simplest of text-based interfaces [RN96]. This theory of the 'medium as social actor' is of direct interest to avatar design because it suggests that minimal cues can elicit social responses. Biocca, Harms and Burgoon maintain that:

> "unlike the physical environment, social communication in virtual environments might be built upon minimal or constrained social cues. Animated characters and even the computer interface itself can generate strong automatic social responses from minimal social cues. Social responses to computer characters for example, are generated even though the user is quite aware that the computer is not an emotional or social agent but a machine" [BHB02, p.5].

They later state that "a fundamental question in mediated social presence is why humans respond automatically and socially to virtual representations of other beings" [BHB02, p.25]. For Biocca and colleagues, the automatic interpretation of humanoid forms and nonverbal behaviour can lead people to attribute a degree of sentience to virtual humans.

This tension between automatic social responses and the rational knowledge that virtual humans are artificial entities represents a fundamental and engaging issue that has been addressed in a selection of studies in different research institutions. Virtual humans present promising avenues for social research because they enable the controlled manipulation of specific visual and behavioural variables. However, before they can be employed for social research the underlying premise of whether they elicit comparable social responses to real humans needs to be tested. Bente and Kramer's study comparing perceptions of video and agent animation, discussed in Section 2.4.3, was designed with this goal in mind. Based on their findings they conclude that computer animations can indeed elicit realistic socio-emotional responses.

The same underlying question was addressed by Pertaub, Slater and Barker in a series of studies on fear of public speaking, a common and debilitating form of social phobia

[PSB01a, PSB01b]. The motivation was to explore whether VEs could in principle be useful for the treatment of phobics; before any exposure therapy treatment programs could be developed, it was first necessary to assess whether virtual audiences could evoke the required anxiety responses. In the main study [PSB01a], 40 participants gave a talk to a virtual audience comprised of eight formally dressed male agents seated in a seminar room. Though participants gave a minimum of two talks, the analysis treated this as a between-groups design, considering only the data from the first talk. Three conditions were compared: a receptive 'positive' audience, a hostile 'negative' audience, and a 'neutral' audience that was static. Agents in the animated conditions had a repertoire of ten behaviours such as smiling, nodding and leaning forward (positive), and yawning, shaking heads and walking out of the room (negative).

The aim was to investigate whether the type of audience would affect speakers' emotional responses, including confidence and anxiety levels. Results indicate that for the positive and neutral audience, participants' post-talk confidence level correlated with their pre-talk levels. Conversely, the negative audience resulted in uniformly higher levels of anxiety irrespective of participants' usual confidence as public speakers. Somatic responses followed a similar pattern, with the negative audience resulting in significantly higher levels of self-assessed physical anxiety responses including sweating, tremors and heart palpitations. A separate report on the qualitative findings from the debriefing sessions confirms the anxiety induced by the negative audience and illustrates that participants responded socially to the agents despite knowing that they were not real people [PSB01b].

Blascovich, Loomis and colleagues address similar questions regarding virtual humans' ability to provoke social responses [Bla02, BLB+01, LBB99]. Their research is aimed at establishing whether IVEs can be used to study social interaction, in particular social influence effects such as conformity, social comparison, facilitation and inhibition. Bailenson et al. [BBBL01] report on a study investigating subjective and behavioural responses to an agent with varying levels of visual and behavioural fidelity. Visual fidelity was manipulated by altering the polygonal mesh and texturing technique so that one version of the agent appeared visibly smoother. Behavioural fidelity was manipulated across five conditions representing increments in the agent's eye gaze realism:

1. eyes closed;
2. eyes open;
3. eyes open and blinking;
4. same as 3, plus constant gaze at participant;

5. same as 4, with added pupil dilation at close range.

The sixth control condition replaced the agent with an inanimate cylinder. Under the guise of a memory task, immersed participants were instructed to approach a static male agent and observe his hair and eye colour as well as what was written on front and back of his shirt. They experienced only one gaze condition, repeating this task five times with a lower-realism avatar and five with a higher-realism agent; the agent's appearance was modified slightly for each study. Tracking data was used to test Argyle and Dean's intimacy equilibrium theory, which predicts an inverse relationship between gaze and interpersonal distance [AD65]. Their analysis indicates that as predicted, both men and women kept a significantly greater distance from the agents in the more 'realistic' gaze conditions (4 and 5). In terms of subjective responses, there was a significant correlation between higher gaze realism and reported social presence among female participants. Overall, the authors conclude that although no participants thought the agent was controlled by a human, they nevertheless clearly did not treat it as a 'mere animation.' The additional interesting finding is that no effect is reported for the agent's appearance. The issue of how visual realism interacts with eye gaze realism is addressed in the experiment presented in Section 5.2.

The above studies suggest that limited visual feedback from virtual humans can affect social responses even in the absence of two-way verbal exchange, and in spite of a rational awareness that these are artificial entities. All of the above results concern behaviour. In terms of visual realism, Nowak and Biocca report on a between-subjects, non-immersive study investigating the role of morphology and agency [NB01]. Agency, defined as the "core volitional or intentional force that drives the actions of an entity" [NB01, p.2], was manipulated by informing participants that they were interacting either with a human or agent. Anthropomorphism (the degree to which the characters resembled a human body) was manipulated by altering the appearance of the character. In the high-anthropomorphism condition participants saw a female face, and in the lower-anthropomorphism condition they saw a more abstract face with only mouth and eyes. A third control condition contained no image.

The interaction consisted of a brief exchange of information regarding a scavenger hunt task. The conversation was bounded and highly artificial, consisting of two speaking turns each; the virtual confederate delivered two pre-recorded audio segments, and participants were required to press a button to indicate the end of each of their turns. Their analysis reveals no effect for agency on either presence, social presence or copresence[4]; while interpreting this as further evidence supporting Reeves and Nass' theory of the 'medium as social actor', the

[4]The measurement instruments will be discussed in greater detail in Sections 2.5.7 and 2.5.9.

authors caution that the results are based on null findings and should be interpreted with caution. Conversely, anthropomorphism is reported as having a significant positive effect on presence, social presence and copresence, with the more anthropomorphic agents outperforming the more abstract characters. While interesting, it is arguable that these results should also be interpreted with caution given the contrived nature of the interaction. Further research is necessary to establish how these results might generalise to sustained interactions.

This subsection has discussed related studies exploring the impact of minimal fidelity on social perceptions of virtual humans. Different aspects of behavioural and visual fidelity have been explored. One important subcomponent of behavioural fidelity that has received less explicit attention is responsiveness. Swinth and Blascovich define this as "the extent to which avatars are capable of responding in socially meaningful and appropriate ways. Interactional realism can range from low to high, with low interactional realism being reflected by a static, non-interactive avatar and high interactional realism being reflected by a dynamic, interactive, and potentially responsive one" [SB02, p.331]. The impact of responsiveness is investigated in the experiment presented in Chapter 4, comparing social responses to agents displaying incremental increases in animation and responsiveness.

2.4.5 Summary of research approach

The general approach taken in the research presented in this thesis has been to explore the impact of minimal fidelity on social responses. Three aspects of fidelity are considered: visual fidelity in terms of appearance, and behavioural fidelity in terms of both animation and responsiveness (Figure 2.6).

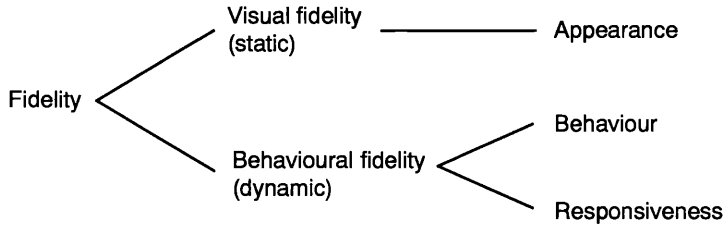


Figure 2.6: Aspects of virtual human fidelity considered in the research presented in this thesis

Bowers, Pycock and O'Brien suggest that "a viable and systematic research strategy for developing useful CVEs is to incrementally add further sophistication to very simple embodiments *as and when* analysis reveals that it is called for in the support of social interaction" [BPO96, p.65]. Isolation of any single behaviour obviously compromises the 'gestalt' of nonverbal expression that characterises face-to-face interaction. However, varying individual dimensions

allows researchers to explore their individual impact on the perceptual and social impact of avatars. This is a logical approach given the need to prioritise which aspects of fidelity might be traded off against available computing resources.

The choice of eye gaze as the primary behaviour for investigation was first prompted by comments from the actors in the virtual rehearsal experiment [SHS+00]. They explained that one of the main barriers to interaction was the inability to look at each other through the avatars. Gaze also seemed a fruitful point of departure given its central role in face-to-face interaction and mediated communication.

CVEs preserve spatial cues, useful in the portrayal of directed attention. For this reason it becomes important to consider the importance of gaze behaviour, as this is the prime indicator of attention. Biocca and Harms [BH02] cite four functions of nonverbal behaviour that are essential in establishing access to other minds in the context of mediated interaction: the communication of attention, emotion, comprehension and behavioural interdependence. Eye gaze contributes directly to each of these four functions, and therefore provides a point of departure for investigating the impact of avatar fidelity on social interaction in CVEs.

In summary, the approach taken in the research presented in this thesis is one that is adopted by many designers of CVEs: exploiting minimum cues to obtain maximum results. This section has focused on what is meant by minimal cues by discussing aspects of fidelity pertinent to avatar design. The final issue to be addressed in this chapter concerns what is meant in this case by 'maximum results.' In the words of Nowak and Biocca, "The purpose of embodiment and agency of virtual humans is to give the user a sense of the other's presence. So a key performance goal of many social virtual environments is to feel as if you are 'there' (telepresence) in the 'company of others' (copresence)" [NB01, p.13]. The following section will discuss the definition and measurement of presence, social presence and copresence.

2.5 Assessing Experiences in CVEs in terms of Presence, Social Presence and Copresence

In assessing the potential of CVEs as a medium of communication it is important not to lose sight of the central goal of giving users a sense that they are 'with' others in a shared environment. Biocca and Harms argue that the goal of the technological effort to produce effective technologies for mediated communication is "the creation of social presence and the improvement of social communication" [BH02, p.12]. The study of social interaction in CVEs draws on interrelated areas of research: presence, social presence and copresence. The goal of presence research is to understand what leads to people's sense of 'being there' in the VE despite

the fact that we know rationally it is not 'real' [Bio97, SUS94, IDRFA00]. Presence is of interest to practitioners in a number of fields including engineering, computer science, psychology, cognitive science, communication and philosophy [DS00, LD97, IFdR01], as well as telecommunication and teleoperation [IFdR01]. Research has been driven both by theoretical and practical concerns [LD97], since a heightened sense of presence is considered essential for effective psychotherapy [HRK+95], for performance in training simulations [Bio97] and for a wide variety of other VE applications. Similarly, research on social presence and copresence aims to understand how to enhance the sense of being with other people in the shared VE; it is of interest for all those applications that involve some form of social interaction, from collaborating with remote human users [BBF+95, SSA+01, Sla99b], to virtual acting rehearsal [SHS+00], to practising public speaking with a virtual audience [PSB01b, PSB01a]. It has been argued that the sense of copresence may be "essential for group collaboration" [TBS+98, p.57].

Limited empirical research has been conducted on the relationship between presence and social presence, or copresence. There exist different schools of thought on how they might interrelate. De Greef and IJsselsteijn argue that:

> "Social presence, or the 'sense of being together', is quite different from physical presence, or the sense of 'being there' in a mediated environment. Although a number of medium manipulations will have a similar effect on both social and physical presence, and a unifying definition has been proposed, the two types of presence can be meaningfully distinguished" [DGI00, p.2]

Thus any correlation may be caused by shared determinants, not by any logical link between the two. Slater et al. similarly argue that though presence and copresence may co-vary, this does not necessarily suggest a causal relationship [SSUS00]. Schroeder claims that "shared VEs often combine a high degree of presence with a high degree of copresence because the sense of being in another place and of being there with another person reinforce each other" [Sch02b, p.5], suggesting a reciprocal link between the two phenomena. Elsewhere, Schroeder argues that the sense of presence and copresence are also likely to reinforce each other [Sch02a]. Heeter takes a different view, defining social presence as a subset of presence, and also as an overall enhancer of presence: "The premise of social presence is simply that if other people are in the virtual world, that is more evidence that the world exists. If they ignore you, you begin to question your own existence" [Hee92, p.266]. Here, a unidirectional causal relationship is proposed between social presence and presence. Heeter emphasises the importance of

responsiveness (both in the virtual humans and objects encountered in the VE) in enhancing the overall sense of 'being there.' The question of how virtual human responsiveness affects social responses and presence is addressed in the experiment presented in Chapter 4 of this thesis. As Schroeder has cautioned, however, there is little empirical evidence for a causal relationship between these constructs [Sch02c]. Sallnäs argues that further investigation into the connection between presence and its social counterparts is needed, since both are believed to be predictors of improved performance but their interrelationship is not known [Sal02]. Table 2.8 summarises a range of views on the relationship between presence and social presence (or copresence).

Table 2.8: Differing views on the possible relationship between presence and social presence

Relationship between Presence and Social Presence	**Comments**	**Researchers**
Unknown	Relationship between presence and social presence or copresence is unknown	Slater et al. [SSUS00], Sallnäs [Sal02]
Indirect	Presence and social presence are distinct constructs that may co-vary because they share determinants	De Greef and IJsselsteijn [DGI00]
Reciprocal	Presence and social presence have a reciprocal relationship, and may reinforce each other	Schroeder [Sch02c]
Causal	Social presence increases presence because a social entity responding to a participant in the VE will enhance the overall sense of presence	Heeter [Hee92]

In addition to the open questions about how presence, social presence and copresence interrelate, there is a lack of consensus on the definition of each of these three terms. Although not the primary focus of this thesis, a discussion of presence is warranted given its probable connection with social presence and copresence. Also, two of the experiments presented in Chapters 4 and 5 were conducted in an immersive setting, making it possible to jointly investigate presence and social responses to avatars. This section will therefore present a range of views on the definition, determinants and measurement of presence. The definition of social presence and copresence is equally problematic, and the fact that the terms are often used interchangeably adds to the confusion. These terms have distinct origins and will be introduced in their original usage. A discussion of how each of these terms has been appropriated in discussions of CVEs will follow. This section will conclude with a statement of how the terms will be applied in the context of this thesis.

2.5.1 Presence

The term *presence* finds its origins in the term *telepresence*, used in the context of teleoperation research to describe the crucial sense of 'being there' in the remote physical environment during remote operation of machines [IFdR01, She92a]. Present conceptualisations of presence vary in their scope. Blascovich broadly defines it as "a psychological state in which the individual perceives himself or herself as existing within an environment" [Bla02, p.129]. This definition applies equally to one's immediate physical surroundings, to a remote physical setting as in teleoperation, and to environments that are computer-generated or simply imagined. Sheridan's definition narrows the discussion of presence to mediated experience, making a further conceptual distinction between telepresence (the sense of being present at a remote teleoperator site) and virtual presence (the sense of being present in a computer-generated environment) [She92a, She96, She92b].

The notion of presence has been applied to a wide range of contexts, ranging from books or films [Zel92], to broadcast media including television and 3D IMAX films, to communications systems including the telephone and VMC [LD97]. Lee and Nass apply the term presence to the full spectrum of mediated experience, claiming that "presence is at the heart of humans' desire to use media to move beyond the limits of body and the sensory channels" [LN01, p.3]. Most recently, presence has received significant attention in the literature on VEs. Biocca contends that VEs differ from other media in the quality and quantity of presence they make possible, and that presence becomes an explicit goal with the advent of immersive technology [Bio97]. Three central issues will be discussed in relation to presence: its definition, its possible internal and external determinants, and measurement approaches.

2.5.1.1 Definitions

Lombard and Ditton's definition of presence as "the illusion of nonmediation" [LD97] has been frequently cited in the literature (see for example [BM02]). They use this definition to unify six separate conceptualisations of presence found in the literature, as summarised in Table 2.9.

Aspects of this classification scheme appear confusing, such as the subtle distinction in the 'transportation' category, between a space being transported to the user and a user being transported to a space. Nevertheless, the overall definition helps to focus the discussion on what it means to be 'present' in the context of a mediated experience. Lombard and Ditton propose two alternative ways in which the illusion of nonmediation can occur. Firstly, a medium can appear transparent; secondly, it can be transformed into something other than a medium, namely a social entity. This second alternative links closely with the notion presented by Reeves and Nass that people can respond to media as social actors even in the presence of minimal cues

Table 2.9: Six conceptualisations of presence united by Lombard and Ditton under the banner of the "illusion of nonmediation" [LD97]

Conceptualisation of Presence	Summary of conceptualisation
Social richness	Draws from Short, Williams and Christie's notion that media richer in informative cues enhance social presence [SWC76].
Realism	The degree to which a medium can convey accurate portrayals of objects, events and people. Divided into *social realism* (the extent to which what is portrayed would be plausible in real life), and *perceptual realism* (the extent to which events appear realistic).
Transportation	The extent to which a person is 'transported' to another place (e.g. through fiction), the extent to which another place is 'transported' to the person's physical environment (e.g. through film), and the extent to which people are transported to a 'shared space' through mediated interaction.
Immersion	The extent to which a person is perceptually or psychologically immersed as a result of substituting real-world stimuli with stimuli from the medium.
Social actor within the medium	The extent to which people respond to social cues presented by people encountered within the medium, even when it is not appropriate to do so (e.g. responding to television characters).
Medium as social actor	The extent to which the medium itself (e.g. a computer program) is responded to as a social entity.

[RN96].

Lombard and Ditton's definition links with the earlier notion of virtual presence proposed by Sheridan [She92a, She96, She92b], emphasising a loss of awareness of one's immediate physical context. Barfield and Weghorst present a similar argument that "presence in a virtual environment necessitates a belief that the participant no longer inhabits the physical space but now occupies the computer-generated virtual environment as 'place"' [BW93, p.702]. In these views, presence seems to be presented as a tension between the physical and mediated environment. Schroeder's definition of presence as "being in another place other than the one you are physically in" [Sch02a, p.289] captures this tension between two alternative environments vying for belief and attention.

Draper, Kaber and Usher point out the limitation of this 'bifurcated' view, stating that presence is more frequently viewed as a scale representing a range in intensity [DKU98]. In the case of VEs, discussions of presence often focus on the degree to which computer-generated sensory cues (visual, auditory and haptic) can replace cues from the physical environment, drawing users into a sense of presence in an environment that is not physically real [She96]. In this context, presence can be viewed as a continuum as opposed to a binary construct. Zeltzer proposes that in real life "we are immersed in a very high bandwidth stream of sensory input, organised by our perceiving systems, and out of this 'bath' of sensation emerges our sense

of being in and of the world" [Zel92, p.128]. He goes on to define presence as "a rough, lumped measure of the number and fidelity of available sensory input and output channels" (Ibid). Zeltzer treats presence as a property of the VE, specifically its capacity to immerse a user in this 'sensory bath.'

Heeter presents a seemingly related definition of presence as "a momentary subjective state of reacting to current impinging sensory stimuli" [Hee01, p.9]. The key difference, however, is that while Zeltzer views it as a property of the medium, Heeter like many others views presence as a subjective state that varies over time. She argues that Lombard and Ditton's definition of presence as the "illusion of nonmediation" [LD97] is restrictive in that presence can also apply to non-mediated experiences. The argument here is that one's sense of presence in the real world can also fluctuate when attention varies between the concrete (perceptual) and the abstract (conceptual). A similar notion is introduced by Slater Usoh and Steed [SUS94], Witmer and Singer [WS98] and Biocca [Bio97], of a third, 'imaginal' environment (akin to daydreaming) that constitutes an additional point of tension between the physical and virtual space. This challenges the assumption that people's awareness must be binary, focusing either on the virtual or physical environment.

Elsewhere, Lombard and Ditton's definition of presence has been further subdivided into spatial and social aspects (see Table 2.10 for a summary). These are referred to respectively as *telepresence* and *social presence* by Biocca and colleagues [BHB02, BBHS01], and as *physical presence* and *social presence* by IJsselsteijn, Freeman and de Ridder [IFdR01]. Heeter similarly subdivides presence into three dimensions: subjective personal presence, social presence, and environmental presence [Hee92]. Subjective personal presence is defined as "a measure of the extent to which and the reasons why you feel like you are in a virtual world" [Hee92, p.262]. This definition is based on sensory perception of simulated stimuli, and it is unclear how this differs from the general sense of presence as defined by other researchers. Social presence refers to the extent to which other human or artificial beings respond to you in the VE, and environmental presence refers to the responsiveness of the VE itself. Heeter postulates that a highly responsive VE might engender a greater sense of presence by virtue of its ability to react to the user's presence, even if its way of reacting violates the logic of the real world. This question relates to Lombard and Ditton's notion of social realism (see 'realism' in Table 2.9), and would require empirical validation.

Researchers vary in their definitions and classifications of presence, resulting in a proliferation of terms (see [DKU98]). Discussions of social presence suffer from a parallel lack of consensus, because the term is often poorly defined and is often used interchangeably with

Table 2.10: Subdivision of the term 'presence' into spatial and social aspects

Aspect of presence	Term used	Researcher
Spatial	Telepresence	• Biocca, Harms and Burgoon [BHB02, BBHS01] • IJsselsteijn, Freeman and De Ridder [IFdR01] • Heeter [Hee92]
	Physical presence	IJsselsteijn, Freeman and De Ridder [IFdR01]
	Environmental presence	Heeter [Hee92]
Social	Social Presence	• Biocca, Harms and Burgoon [BHB02, BBHS01] • IJsselsteijn, Freeman and De Ridder [IFdR01] • Heeter [Hee92]

copresence (see Section 2.5.4 for a discussion of social presence and copresence).

2.5.1.2 Internal and external determinants

Discussions of presence also target the question of how the sense of presence is created and destroyed. IJsselsteijn et al. argue that "although research into presence is still at an early stage of development, there is a consensus that presence has multiple determinants" [IDRFA00, p.1]. Several possible determinants of presence have been proposed, encompassing both technical characteristics of the media system, individual characteristics of users, and task and context variables. Slater, Usoh and Steed suggest that determinants could be categorised as internal (endogenous) factors relating to the individual, or external (exogenous) factors relating to the medium [SU93, SUS94].

Internal variables, also referred to as person variables [BW93] or user characteristics [IDRFA00], have received less research attention compared to external media form variables. Slater and Usoh investigated the impact of individuals' dominant representation system (according to the visual, auditory and kinesthetic classification used in neurolinguistic programming) [SU93]. Freeman, Lessiter and IJsselsteijn propose additional user characteristics that may affect presence: these include prior experience with media (and resulting expectations), mood, personality, age and gender, as well as perceptual, motor and cognitive abilities [FLI01]. Given the possible impact of individual characteristics on presence and possibly also social presence, a number of explanatory variables were collected in the studies reported in Chapters 4 and 5.

A number of possible external determinants of presence have also been proposed. Free-

man, IJsselsteijn and colleagues list four classes of presence determinants identified in the literature [FLI01, IDRFA00]:

1. the extent and fidelity of sensory information;
2. the match between the sensors and the display;
3. content factors;
4. user characteristics.

The first two are classified as media form variables, relating to properties of the system. As illustrated in Table 2.11, the *content factors* category is broad, and is captured with greater granularity in the earlier categorisation scheme of Slater, Usoh and Steed [SUS94][5].

Slater and Steed also propose a number of factors that undermine presence, causing breaks in presence (BIPs) or 'transitions to real' where people's attentional focus is suddenly drawn out of the VE to their physical surroundings [SS00]. These factors can be either external (sensory information from the physical world intruding or contradicting the VE), or internal (internal inconsistencies in the VE).

Given the range of media factors that may impact on the sense of presence, some authors have made a point of conceptually distinguishing between the presence experience itself and its possible determinants. The term *immersion* is used to describe the extent to which objective characteristics of the technology can provide a surrounding environment by replacing sensory stimuli from the physical world [SW97, SU95, Sla99a, DKU98][6]. A parallel distinction is made by Biocca and others with regard to the factors that enable social presence, and the experience itself (see Section 2.5.6).

2.5.1.3 Measurement approaches

A number of measurement approaches have been proposed, some of which are summarised in Table 2.12. These can be classified according to the time measurement is taken (during or after the experience), and the type of data gathered (subjective or objective).

Presence is frequently referred to as a subjective experience [IDRFA00, She92b, WS98]; unsurprisingly, presence research has relied extensively on subjective reporting, most commonly on the use of post-experiment questionnaires designed to evaluate people's sense of

[5]Note that the same original sources are treated in both categorisation schemes.

[6]The term 'immersion' is confusingly also used by Witmer and Singer to describe the sense of "perceiving oneself as a part of the VE stimulus flow" [WS98]. Though they claim this sense contributes to presence, they do not clearly state how it differs from it. They further claim that VEs are the only medium enabling immersion, but do not clarify if presence is possible in non-immersive media.

Table 2.11: Classification of possible external determinants of presence

Freeman, IJsselsteijn and colleagues [FLI01, IDRFA00]	**Slater, Usoh & Steed** [SUS94]	**Others**
Extent and fidelity of sensory information	Transparent presentation of high-quality, high-resolution sensory information	Fidelity and extent of sensory information (top priority): Sheridan [She92b]
		• Transparent interface • high resolution • Large field of view Held and Durlach [HD92]
		• Presence (number and fidelity of sensory channels simulated by the VE) • Autonomy (the ability of the VE to simulate real-world interactions): Zeltzer [Zel92]
		• Display fidelity (spatial and contrast resolution) • Environmental stability Barfield and Weghorst [BW93]
	Consistency	Consistency of information across sensory channels: Held and Durlach [HD92]
Match between the sensors and the display	Straightforward relation between actions and effects in the VE	Degree of sensor control in the environment: Sheridan [She92b]
		• Sensorimotor consistency between real and virtual movement • Wide range of sensorimotor interactions (the ability to navigate freely) Held and Durlach [HD92]
Content factors	Interactivity	Interaction (the ability of the VE to respond in real time to the user's input): Zeltzer [Zel92]
		Interactive fidelity: Barfield and Weghorst [BW93]
	Self-representation in the VE consistent with real-world appearance	Identification with virtual self through similarity in visual appearance: Held and Durlach [HD92]
		• Task variables • Context variables Barfield and Weghorst [BW93]

Table 2.12: Proposed measurement approaches

	During experience	Post-experience
Subjective	Hand-held slider	Questionnaires
	Breaks in presence (BIPs)	Interviews and Focus groups
Objective	Psychophysiological monitoring	
	Observation of behaviour	

'being there' in the mediated environment. Subjective questionnaire measures can combine different approaches [DGI00], including semantic differential techniques using scales anchored to opposing descriptors, as in [WS98]. Alternatively, Likert scales have been used to measure the degree of agreement or disagreement with a set of statements, as in [SSU93].

Two presence questionnaires have received significant attention in the literature: the Witmer and Singer presence questionnaire (PQ) [WS98], and the Slater-Usoh-Steed questionnaire (SUS) [SUS94]. Witmer and Singer's PQ was developed to elicit subjective presence responses to experiences in IVEs, with a particular focus on investigating the impact of four possible contributing factors to presence: control, sensory factors, distraction and realism. The problem, as discussed by Slater [Sla99a], is that the questionnaire confounds measures of individual differences and properties of the VE, making it impossible to separate them. In addition, although they clearly define presence as "the subjective experience of being in one place or environment, even when one is physically situated in another" [WS98, p.560], their questionnaire contains no items that directly measure this construct.

Slater, Usoh and Steed's SUS questionnaire is designed to measure the sense of 'being there' in the VE, as well as two additional aspects central to Slater's definition of presence: the extent to which the VE is experienced as the dominant reality, and the sense of having visited a place as opposed having simply viewed computer-generated images [Sla99a]. This sense of place is particularly central to the experience of presence in VEs.

Usoh et al. report on a study designed to test the ability of both the PQ and the SUS questionnaires to distinguish between subjective presence responses to a real-world environment and its corresponding immersive virtual model [UCAS00]. They report that PQ showed no difference between the real and virtual environments, while SUS showed a statistically significant difference.

A tradeoff is involved in using post-experience questionnaires. One significant limitation is that subjective reporting only captures post-hoc rationalisations of the experience. This is problematic not only because of demand characteristics [DKU98], but also because of the po-

tential pitfalls of inaccurate recall [FAPI99]. Freeman et al. have pointed out that post-test presence ratings are unstable, particularly in the case of naive subjects who lack a lexicon for understanding and describing presence [FAPI99].

Slater has similarly argued for a move away from questionnaires in the measurement of presence [Sla04]. In a recent study, a questionnaire referring to a fictitious construct called "colourfulness of an experience" was administered to 74 respondents. Reported findings indicate an association between 'colourfulness' and a number of equally arbitrary variables including how late respondents had woken up that day. Slater cautions that questionnaire responses can yield statistically significant but ultimately 'meaningless' results because rather than reflecting how respondents would ordinarily describe their experience, the arbitrary response measure is called into being by the questionnaire.

It has further been argued that far from being a stable constant throughout the mediated experience, presence may vary over time [She92b, KB97, BBHS01, IFdR01]. In Heeter's words, "Presence is a momentary subjective state of reacting to current impinging sensory stimuli" [Hee01, p.9]. It is challenging to capture these fluctuations in a post-test rating, however an effort was made in two of the experiments presented in Chapters 4 and 5 of this thesis to address these temporal fluctuations by asking participants about their *initial* responses versus responses *over the course of the whole experience.*

Alternative approaches have been explored to address the temporal fluctuations in presence. Slater and Steed propose a "breaks in presence" (BIPs) approach, where participants are asked to signal each time they transition to a state of awareness of their physical surroundings [SS00]. This method presumes a binary possibility whereby people are either present in the VE or in the physical environment. By the authors' own admission, this method fails to capture presence in a third 'imaginal' location. Nevertheless, its advantage is that Slater and Steed's findings suggest a strong positive correlation between questionnaire-based presence and presence as estimated from the number of BIPs reported. Freeman et al. have explored an alternative method, attempting to capture temporal variations in presence through continuous reporting using a hand-held slider [FAPI99]. The significant drawback of both the BIPs and slider approaches is their intrusiveness; by requiring participants to continually report on their experience, these methods introduce additional cognitive load and also potentially interfere with the phenomenon of interest, the presence experience itself.

Objective approaches have been investigated to address the limitations of both continuous and post-test subjective ratings. Their advantage is that they do not require conscious attention or control and are therefore less cognitively intrusive. Held and Durlach suggest observing

involuntary behaviours as indicators of presence, such as ducking or blinking when a virtual object approaches one's head [HD92]. Other possibilities include 'socially conditioned' responses such as reaching for objects or attempting to shake the hand of a virtual person [She92b]. Noting the potential instability of subjective ratings found in their earlier work [FAPI99], Freeman et al. explored the possibility of employing behavioural responses to measure presence [FAM+00]. The rationale behind using this approach is that people will respond in 'behaviourally realistic'[7] ways to a mediated experience to the degree that it accurately simulates an equivalent real-world experience. They report on a study investigating the impact of monoscopic and stereoscopic (3D) displays on subjective presence ratings and postural responses to vection (the illusion of observer motion provoked by moving displays). Postural shifts in response to video images of a speeding car were measured using a magnetic position tracker. Though the stereoscopic image resulted in both increased lateral movements and subjective presence ratings, no significant relationship was found between the two. The authors therefore caution against direct substitution of postural responses with self-reporting, though arguing that they may be usefully employed in the evaluation of displays.

Following a similar rationale, Meehan et al. explored the possibility of using psychophysiological measures including skin temperature, heart rate, and electrodermal activity (EDA) to measure presence [MIWB01]. Here presence is equated with the success of IVEs in recreating real world experience; the underlying assumption is that if present, a person should exhibit similar psychophysiological responses to an analogous real-world situation. Their study investigated participants' responses when entering a virtual 'pit' room containing a steep drop-off to the floor below. Findings suggest a steep rise in mean level of heart rate and EDA, indicating a marked increase in arousal, as would be expected in an equivalent real-world setting.

The appeal of both these objective approaches is that participants tend to be unaware of both postural shifts and psychophysiological responses, so that responses are not as likely to be affected by demand characteristics and other forms of higher-level cognitive processing. In addition, Freeman et al. argue that postural shifts vary in intensity, providing a graded as opposed to binary measure of the objective response [FAM+00]. The drawback, however, is that both approaches require specific, and in the case of the pit room, extreme stimuli to elicit a measurable objective response. Though Barfield and Weghorst suggest that subjective measures may eventually be replaced by more robust, objective measures [BW93], content-dependency may

[7]Freeman et al. refer to this as the 'behavioural realism' approach. However, in the context of this thesis, 'behavioural realism' refers specifically to the behavioural fidelity of the avatar, and not to the participant's behaviour as an objective measure of presence in a VE.

provide a significant barrier to the use of objective measures. In the words of Freeman, Lessiter and IJsselsteijn, "content-dependency makes the development of a general behavioural metric unlikely" [FLI01, p.8]. Given the limitations of both subjective and objective measurement approaches, Freeman et al. have proposed the the parallel exploration of objective and refined subjective measurement approaches. In particular they propose the use of focus groups to derive improved terminology for rating scales. An aggregate approach combining various measures may be more effective, particularly considering the potentially complex structure of presence. As Slater, Usoh and Steed suggest, presence may consist of two levels: the surface level, which can be consciously articulated, and a deeper level that "influences behaviour in a basic way" [SUS94, p.142] and may be better captured by objective means.

A range of views on the definition, determinants and measurement of presence have been summarised. A discussion of social presence and copresence will follow, beginning with the original definitions of each term.

2.5.2 Short, Williams and Christie's original definition of social presence

The term *social presence* as applied to mediated communication can be traced to Short, Williams and Christie's 1976 publication of *The Social Psychology of Telecommunications* [SWC76]. At the time of writing, a variety of new visual telecommunication systems were becoming available. In comparison with one-to-many mass communication media, person-to-person telecommunications media had received relatively little research attention. Emerging research into 'technology assessment' was concerned with investigating telecommunications media in terms of their emergent uses, as well as their impact on social interaction. In particular, the authors were interested in the circumstances under which people would willingly substitute face-to-face with mediated interaction. They were therefore concerned with the medium's *perceived* capacity to transmit the rich visual cues available in face-to-face communication:

> "We conceive of Social Presence not as an objective quality of the medium, though it must surely be dependent on the medium's objective qualities, but as a subjective quality of the medium. We believe that this is a more useful way of looking at Social Presence than trying to define it objectively" [SWC76, p.66].

The focus on subjective perceptions is partly motivated by the authors' belief that the adoption or rejection of a new medium hinges not on its intrinsic value, but on individuals' opinions about its ability to meet their telecommunications needs. The research emphasis is therefore on preference and perceived effectiveness, rather than on objective properties of the medium.

Short, Williams and Christie's focus is on dyadic or small group interaction, with a particular emphasis on how the transmission of nonverbal cues is seen to be affected by the medium used. Their measurement approach is based on a semantic differential technique that rates media on a series of seven-point, bipolar scales including unsociable-sociable, insensitive-sensitive, cold-warm, and impersonal-personal. High Social Presence[8] equates with the medium being seen as sociable, sensitive, warm and personal. In this view, Social Presence is expected to vary significantly between media and to be highest in face-to-face interaction. A series of studies are presented, suggesting that Social Presence has sufficient granularity to distinguish between face-to-face meetings and mediated group conferences, as well between different versions of the same medium. The pattern of findings supports their hypothesis that "Non-verbally rich media are more favourably evaluated than non-verbally poor media particularly on the Social Presence scales" [SWC76, p.115].

Short, Williams and Christie relate their definition of Social Presence to Argyle and Dean's concept of *intimacy*. This concerns the level of attraction people feel for each other, reflected in nonverbal displays of approach and avoidance [AD65]. A variety of nonverbal behaviours come into play including gaze and proxemics (spacing behaviour), as discussed in Section 2.3.3. Short, Williams and Christie posit that a higher level of Social Presence should increase the overall level of intimacy possible through a medium.

In summary, the original definition of Social Presence concerns a medium's *perceived* capacity to transmit the cues available in face-to-face communication. A discussion of the original definition of copresence follows.

2.5.3 Goffman's original definition of copresence

The term copresence was originally coined by Erving Goffman in the context of sociological research on human behaviour in public places [Gof63]. In contrast to Short, Williams and Christie, Goffman's interest lies not in mediated communication (or in what he terms 'disembodied' forms of contact such as messages and letters), but in the social norms governing face-to-face interaction. Particular emphasis is placed on the role of physical proximity: "Interaction (that is, face-to-face interaction) may be roughly defined as the reciprocal influence of individuals when in one another's immediate physical presence" [Gof63, p.26].

The original definition of copresence is therefore firmly rooted in discussions of embodiment and physical space. Copresence applies to two distinct levels of interaction. Firstly, *unfocused* interaction, where people are physically proximate and therefore "uniquely accessible, available and subject to one another" [Gof63, p.22]; secondly, *focused* interaction, where peo-

[8] Short, Williams and Christie's original capitalised form will be used when referring to their definition.

ple engage directly in conversation with one another. Copresence is not limited to situations of actual engagement in conversation (focused interaction), but also encompasses potential availability to interaction by virtue of being close to another person (unfocused interaction). In both cases, it implies a form of active and controlled alertness to the presence of others: "Persons must sense that they are close enough to be perceived in whatever they are doing, including their experiencing of others, and close enough to be perceived in this sense of being perceived" [Gof63, p.17].

Elsewhere, Goffman discusses the theatrical norms governing public behaviour, using such dramaturgical metaphors as 'back' and 'front' regions, audience, performance and backstage behaviours [Gof71]. This notion of theatrical performance is central to the concept of copresence, which concerns not only the individual's position in relation to others but also an active form of responsiveness to the environment. This is coupled with a social obligation to appropriately control one's appearance and behaviour:

> "The disciplined ordering of personal front is one way, then, in which the individual is obliged to express his aliveness to those about him. Another means is the readiness with which he attends to new stimuli in the situation, and the alacrity with which he responds to them with body movements" [Gof63, p.28].

Note that Goffman's definition of presence as form of self-monitoring of *performance* in the environment contrasts with its usage in the literature on VEs, where it signifies a psychological *response* to the sensory stimuli in the environment. Goffman's notion of presence relates more closely to the concepts of fidelity in virtual humans, in terms of believable appearance and behaviour.

Though Goffman's discussion of the full conditions of copresence concerns direct interaction in physical space, Giddens argues that "mediated contacts that permit some of the intimacies of copresence are made possible in the modern era by electronic communications" [Gid84, p.77]. It is unsurprising, therefore, that discussions of interaction in CVEs have appropriated this term.

2.5.4 Appropriation of the original terms in the literature on CVEs

The terms *social presence* and *copresence* have received increasing attention in the literature on avatars, agents and CVEs in recent years [SSA+01]. Given the inherent spatial possibilities of CVEs as a communications medium, it seems logical that researchers should have appropriated both terms. One the one hand, Short, Williams and Christie's notion of Social Presence emphasises mediated communication and the transmission of nonverbal cues; on the other,

Goffman's concept of copresence concerns embodied interaction in physical space. These definitions therefore reflect complementary aspects of CVEs as a communications medium (Table 2.13).

Table 2.13: Original definitions of social presence and copresence reflecting complementary aspects of CVEs as a communications medium

Social Presence	Copresence
• Mediated communication • Transmission of nonverbal cues	• Interaction in physical space • Embodiment

The problem is that the appropriation of these terms has sparked a departure from the original definitions, partly accounting for the terminological confusion witnessed in current discussions of social interaction in CVEs. The terms social presence and copresence are often used interchangeably in the literature, along with terms including togetherness [AV02] and shared presence [RB02].

The terms are sometimes misapplied. For instance, in [GL01] the term copresence is referred to in the title of the publication, but is used as an arbitrary name given to an experimental condition, with no discussion devoted to the experience of being in a shared virtual space. The terms are often under-defined; many researchers provide cursory and circular definitions, or sidestep the problem of definition altogether. In the words of Biocca and colleagues, "We and others have sometimes defined social presence as the 'sense of being with another' or the 'sense of being together' in a virtual environment. While this can be useful as a short-hand communication, it is inadequate as a definition. It merely restates the idea of social presence in different words without significant concept explication" [BBHS01, p.11]. In cases where care is taken to define the terms there is the added problem of inconsistency, where researchers' definitions can vary between publications (compare for instances the conflicting definitions of social presence in [Bla02] and [SB02]). In summary, there are at least four problems with current usage of the terms:

1. confounding of the terms *social presence* and *copresence*;
2. misapplication;
3. under-definition;
4. inconsistency.

It is probable that given the complexity of the ongoing presence debate, it will be some time before any consensus is reached on the explication and measurement of social presence and copresence in CVEs. The remainder of this section will focus on the appropriation of the terms in the literature on virtual humans and CVEs. It will begin with a discussion of copresence, followed by a discussion of social presence. An outline of attempts to conceptually distinguish these terms will follow. The section will conclude with an explanation of how the terms will be used in this thesis.

2.5.5 Appropriation of the term 'copresence'

Zhao's usage of copresence stays close to Goffman's original definition in its emphasis on synchronous, embodied interaction [Zha01], while extending its application to mediated interaction. A taxonomy of copresence is proposed in an attempt to give clearer definition to a nebulous term: "The term 'copresence' has recently been appearing in the presence literature with increasing frequency, but, like the concept of presence at its nascent stage, the meaning of this term is yet to be fully explicated" [Zha01, p.1]. Here, the conditions of copresence are classified according to whether proximity is 'physical' or 'electronic', and whether the conversation partner is present 'in person' or 'in simulation'.

The resulting categories in the taxonomy are given similar and confusing labels including *corporeal copresence, corporeal telecopresence, virtual copresence* and *virtual telecopresence*. The notion of a 'simulated other' introduces further confusion when no clear distinction is made between agents and avatars, and when the definition of virtual telecopresence is given as interaction "with a computer program that simulates human responses," [Zha01, p. 5] without specifying whether or not a real person is actually being represented. In short, no consideration is given to the concept of agency. While copresence is defined as "an individual's subjective experience of being together with other people" [Zha01, p.8], the classification focuses on the *conditions* of mediation, as opposed to people's *experience* in different mediated contexts. This approach also does little to explain how the *sense* of copresence might be affected by the various categories of dyadic interaction outlined.

Swinth and Blascovich underline the drawbacks of limiting definitions to descriptions of the medium:

> "Defining social presence or copresence in terms of either a particular technology or a particular interactional context is extremely limiting in that as the technology or social context changes, the constructs and theory lose all utility. Thus, social presence and copresence may better be viewed as general constructs that operate

across interactional settings" [SB02, p.318].

In contrast to Zhao, Schroeder and colleagues define copresence in terms of subjective experience rather than conditions of mediation. Here, copresence is defined as "the subjective sense of being together or being co-located with another person in a computer-generated environment" [SSA+01, p.786], a definition argued to be consistent with the definitions in [SSUS00, SSS+99, WAS+00]. This definition focuses the scope of copresence, limiting its application to CVEs and excluding other media such as videoconferences, telephone conversations and text-based interactions. Elsewhere, Schroeder critiques Zhao for lumping CVEs in the same category as these other media that fail to meet the required criteria of offering the experience of 'being there together' [Sch02a]. He also explicitly rejects the notion of 'social presence' as being too restrictive, because it concerns the medium instead of people's experience of using it [Sch02a]. Schroeder's definition of copresence therefore emphasises the subjective experience of being in a shared, computer-generated space, an experience that is likely to be different from other media: "There is a fundamental difference between visual channels in VR and in existing technologies such as videophones or computer-mediated communication that incorporates 2D graphics, given that VR provides a manipulable and navigable space in which communication takes place" [Sch96, p.146].

Schroeder argues there is a close relationship between presence and copresence, though copresence is a potentially more problematic construct requiring further definition, because it is more "diffuse" [Sch02a, p.275]. Copresence covers a range of issues including the ways others are experienced and engaged with, as well as what they are able to jointly do and achieve together. Particular attention is given to the notion of joint action in the CVE: "Copresence is more about what participants do together rather than being aware of each other's presence" (Ibid., p.291). Though not stated explicitly, it appears that this notion of copresence applies to what Goffman terms *focused interaction*, rather than *unfocused interaction* where there is merely a potential for direct interaction due to physical proximity.

Durlach and Slater also place a central emphasis on joint action [DS00]. They posit that the sense of 'togetherness' is engendered through each individual's sense of presence in the VE, combined with communication among participants. The sense of communication derives from verbal interaction and visual evidence of the outcomes of collaboration, where "alterations in the environment are clearly perceived by the other participants" [DS00, p.216]. Durlach and Slater argue that sense of togetherness might be strongest in cases of haptic collaboration, since touch, unlike vision and audition, is not a 'distance sense' and usually implies close physical proximity. Elsewhere, Slater uses the term copresence, defined consistently as "the sense of

being and acting with others in a virtual place" [SSUS00, p.38]. Again, the emphasis, like Schroeder's, is on the concept of shared action in CVEs.

2.5.6 Appropriation of the term 'social presence'

As with copresence, there are a number cases in which social presence is referred to without being concisely defined. Cursory definitions are often provided, such the "sense of being together" [DGI00, p.2]. The definition given by Sallnäs is slightly more informative: "The feeling of being present with another person at a remote location" [Sal02, p.172]; this captures the notion of media bridging the gap between physically distant people. As with copresence, however, the problem is that the concept is too often under-defined, with researchers using "vague, overly broad, or circular definitions of social presence" [BHB02, p.20]. A number of more detailed definitions of social presence have been proposed. The definitions cover different themes, some of which are discussed below.

2.5.6.1 Social presence as the illusion of nonmediation

Lombard and Ditton's definition of presence as the illusion of nonmediation [LD97] has a social subcomponent (see Table 2.9). Becker and Mark explicitly fuse this definition with Short, Williams and Christie's [SWC76]:

> "The degree of social presence is determined by how a number of such non-verbal cues are conveyed by the medium, and influence how present or distant one feels from another person. A high degree of presence suggests the illusion that one is directly interacting with another, and the medium becomes less apparent" [BM02, p.29].

The argument is that the more nonverbal cues are transmitted, the more transparent the medium becomes. Although not explicitly stated, it appears that this definition of social presence focuses on the individual's experience of interaction as opposed to perceptions of the medium.

2.5.6.2 Social presence as access to another intelligence

In the broadest sense, Biocca and colleagues define social presence as the sense of "being together with another" [BBHS01, p.2], although cautioning that further definition is required to address the conflation of an amorphous collection of variables in present research. As in the case of presence, they point out that "there is as yet no widely accepted measure of social presence" [BBHS01, p.6]. While acknowledging that social presence applies equally to face-to-face interaction (as in [Hee01]) and that the terms *mediated social presence* or *social telepresence* would more aptly capture the phenomenon of interest, they adopt the convention

of applying the term exclusively to mediated interactions (as in [Sal02]). Like Short, Williams and Christie, Biocca and colleagues aim to compare media for their degree of social presence [BHB02, BBHS01, BH02]. They argue, however, that a measurement instrument should parallel the majority of presence questionnaires by capturing social presence responses to the interaction *experience*, as opposed to the *medium*.

Elsewhere, Biocca defines social presence as the "degree to which a user feels access to the intelligence, intentions and sensory impressions of another" [Bio97, p.22]. Social presence is assessed according to a medium's capacity to offer a level of interaction capable of sustaining a psychological connection to the intelligent other (whether human or artificial). Lee and Nass similarly define social presence as "the sense that other intelligent beings co-exist and interact with you" [LN01, p.4]. Like Biocca, they also include artificial intelligence (robots and interface agents) in this category.

2.5.6.3 Social presence as a subjective experience

One critique of Short, Williams and Christie's original definition of Social Presence is that it implies a stable response to a medium. The stability of a medium's properties does not, however, translate into a stable experience. Biocca and colleagues argue that "like presence, social presence is a phenomenal state varying during the course of interaction" [BBHS01, p.11]. Blascovich suggests that social presence can fluctuate in two ways: firstly, people can drift in and out of awareness of their immediate surroundings, and secondly the behaviour of an avatar may suddenly appear inappropriate, resulting in a momentary decline in social presence [Bla02]. These examples seem to be taking into account both definitions of social presence: firstly the illusion of nonmediation, where awareness of the medium can fluctuate over time, and secondly access to another intelligence, where seemingly inappropriate behaviour can lead to a momentary lapse in the belief that one is interacting with an intelligent, sentient entity.

Heeter likens the temporal variations in social presence to presence, that also fluctuates over time and partly hinges on moderating variables:

> "Like the general concept of presence, experienced social presence is tied to a specific experience at a particular point in time. Also like the general concept of presence, experienced social presence is contextual, dependent upon the history of the relationship, the communication content of the interaction, and the communication characteristics of the medium used in the interaction" [Hee01, p.10-11].

Heeter subdivides social presence into two dimensions: "*Experienced* social presence is the particular feeling of connectedness experienced by a person during a specific use of tech-

nology for telerelating.... *Expected* social presence is the potential of a communication medium to facilitate meaningful social interaction" [Hee01, p.10-11]. In this way social presence as a subjective experience of interaction is reconciled with the objective properties of the communications medium.

Kaushik et al. similarly define social presence as a two-dimensional construct "with structural affordances of the medium as one dimension, and the experiential aspects of social presence as the other dimensions" [KDKO02, p.402]. In other words, they account both for the perceived properties of the *medium* and for the resulting *experiences* of users. The experiential dimension of social presence is measured in terms of perceived interactivity, subdivided into four components: awareness of self, awareness of others, ease of communicating with interaction partners, and overall evaluations of the experience.

2.5.6.4 Social presence as a multi-level response

While the majority of definitions discussed above apply to all forms of mediated interaction, that of Bailenson, Blascovich and colleagues applies primarily to CVEs [BBBL01, Bla02, BLB+01, SB02]. They present a model where social influence increases as a positive function of social presence. Here, social presence is defined as "a psychological state in which the individual perceives himself or herself as existing within an interpersonal environment" [BLB+01]. Social presence varies as a positive function of two aspects of believability: agency and behavioural realism (see Sections 2.3.1 and 2.4.3 for Blascovich and colleagues' definitions of these terms). In their threshold model of social influence [BLB+01], it is hypothesised that in order to be influenced by a virtual human, a certain level of social presence is necessary. The logic is that people will be more forgiving of low behavioural realism provided they believe avatars represent real people (high agency); conversely, expectations of behavioural realism will be very high for agents, since they represent low agency.

The threshold of social influence varies according to the response system being investigated. The argument is that the threshold for social influence is lower for those lower-level responses including involuntary actions such as socially conditioned and reflex responses. Blascovich cautions that such responses could not be captured using self-report measures, which relate to high-level response systems. One suggested approach to studying low-level responses involves behavioural observation, such as in Bailenson et al.'s proxemics study investigating interpersonal distance from an agent representing different levels of gaze realism [BBBL01] (see Section 2.4.4).

2.5.7 Theoretical distinctions between social presence and copresence

In numerous cases, the use of the term *social presence* or *copresence* reflects a matter of preference. Some researchers, however, have made a deliberate point of conceptually distinguishing between these terms. IJsselsteijn, Freeman and de Ridder separate presence into two categories: physical and social, where "the physical category refers to the sense of being physically located in mediated space, and whereas the social category refers to the feeling of being *together*, of social interaction with a virtual or remotely located communication partner" [IFdR01, p.181]. Here, copresence represents the intersection of the two categories, as "the sense of being together in a shared space, combining significant characteristics of both physical and social presence" (Ibid., p.181-2). A number of media examples are placed in each category, as illustrated in Figure 2.7.

This definition of copresence is constructive because it distinguishes between the sense of being together (afforded by a variety of media such as telephone and online chats), and the explicit sense of being together in a shared space. Some confusion is created by including asynchronous communications media such as e-mail and letters in the social presence category. The inclusion of videoconferencing and videophone in the same category as SVEs appears to contradict the definition of copresence as an inherently spatial construct; unlike SVEs, the majority of VMC systems fail to represent spatial relationships among users and each user's physical 3D context is lost.

Heeter [Hee01] further adapts this classification to include "real World" and "face-to-Face" in the scheme, as in Figure 2.7. Although this is useful in reiterating the spatial aspect of copresence, it blurs the categorisation by including real-world interaction in a discussion of mediated communication.

An additional drawback of both the above schema is that although presence, social presence and copresence are defined in terms of subjective experience, the classification is based on properties of the medium. Like Zhao's taxonomy [Zha01], it focuses on objective categories, and not on subjective perceptions of mediated experience. Three examples of distinctions between social presence and copresence that account for subjective experiences are discussed below.

In a further elaboration of the 'Networked Minds' theory of social presence, Biocca and Harms subdivide social presence into three levels: copresence with the embodied other, psychobehavioural accessibility of the other, and intersubjective social presence [BH02]. At the first and lowest level, copresence[9] is defined as a necessary but insufficient requirement for the

[9]Biocca and Harms use the hyphenated form 'co-presence', however since this does not imply a distinct definition

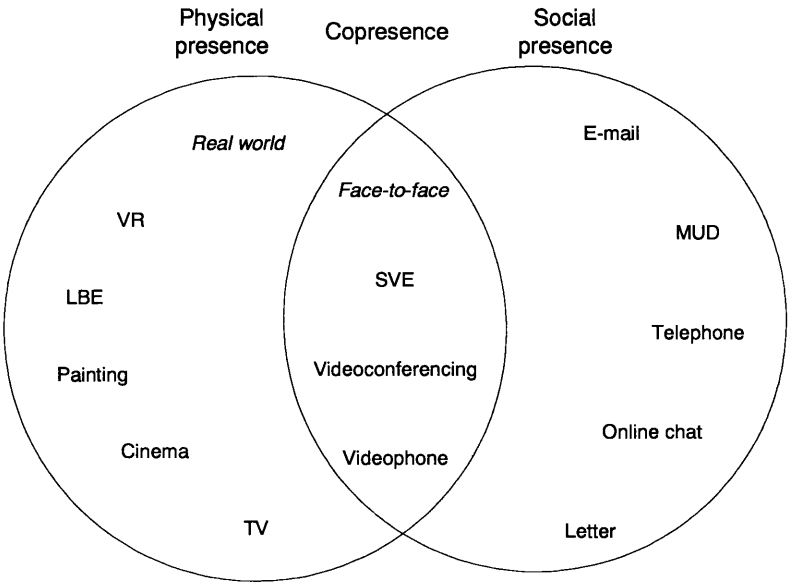


Figure 2.7: Classification of Physical presence, copresence and social presence, from IJsselsteijn, Freeman and de Ridder [IFdR01]. Italicised terms are added by Heeter [Hee01]. (Abbreviations: VR= virtual reality, LBE= location-based entertainment, SVEs= shared virtual environments, MUDs = multi-user dungeons.)

sense of social presence. Here, copresence seems to equate with co-location in that it concerns the "detection and awareness of the copresence of the other's mediated body" [BH02, p.16]. This definition faithfully mirrors Goffman's original definition in its emphasis on the role of the body in communicating mutual awareness in physical space [Gof63]. However, where Goffman's notion of copresence encompassed both potential (unfocused) and actual (focused) interaction, for Biocca and Harms it is limited to the simple subjective assessment of being in the presence of a sentient entity. They propose a copresence threshold where "automatically and without effort, a thing, technology, is suddenly perceived as somehow *being*, a mediated other" [BH02, p.17]. They go on to state that "unlike the physical environment, mediated others can quickly appear or disappear, so the threshold is quickly and frequently crossed in an average day" (Ibid., p.18).

The second level in their model relates to the definition of social presence, proposed in [Bio97] as access to another intelligence. The third and final level of social presence concerns intersubjectivity, the degree of correlation between a person's own sense of social presence and their level of social presence as perceived by their interaction partner. This third level presents

of the term, the non-hyphenated 'copresence' will be used in in this thesis.

social presence as a reciprocal phenomenon (see Table 2.14).

Biocca and Harms' definition attempts to account for the original definitions of both social presence and copresence by combining them in a model that encompasses subjective perceptions, temporal variations, and different levels of engagement. However, given that their theory of social presence encompasses all mediated interaction, no account is provided of how shared VEs might offer a distinct way of experiencing embodied interaction in space. Also, their argument that the copresence threshold can be crossed many times and in a straightforward way raises questions about whether there is a threshold for social presence as a whole, and whether that too can be crossed repeatedly. Although they claim that the overall sense of social presence varies in intensity during the course of the experience, are there factors that might irrevocably damage it? This issue is addressed in the experiment described in Chapter 4 of this thesis, investigating how perceptions of humanoid agents vary during the course of an interaction, and those factors that might permanently damage their credibility as social entities.

Similar to Biocca and Harms, Swinth and Blascovich also propose three different levels to social interaction in mediated environments [SB02] (see Table 2.14). The first level concerns an awareness of social entities in what they call the 'interpersonal environment'. The second level involves understanding whether or not these social entities are human. The third and final level involves determining whether the entities are simply there, or whether they are capable of interaction. There are some parallels with Biocca and Harms' subdivision of social presence into mere copresence (level 1) and access to another's intelligence (level 2). The third level echoes Goffman's distinction between unfocused and focused interaction, in that the presence of other embodied social entities does not necessarily translate into direct interaction.

Table 2.14: Levels of social interaction proposed by Biocca and Harms, and Swinth and Blascovich

	Biocca & Harms 2002 [BH02]	**Swinth & Blascovich 2002** [SB02]
Level 1	Awareness of others mediated body (copresence)	Awareness of social entities in the interpersonal environment
Level 2	Access to another intelligence	Understanding whether the social entities are human
Level 3	Intersubjectivity	Understanding whether social entities are capable of interaction

Where Biocca and Harms treat copresence as a subset of social presence, Swinth and Blascovich view is as a related but distinct construct. Where social presence concerns the external cues provided by the medium, copresence concerns the subjective "perception and feeling that

others are co-situated within an interpersonal environment" [SB02, p.319]. This distinction is problematic because it contradicts the definition of social presence proposed elsewhere by Blascovich as a psychological state of awareness of being in an interpersonal environment [Bla02]. Here it is defined as the "actual, imagined or implied presence of others" [SB02, p.318], where *actual* means physical and *implied* means mediated. It is difficult to see how 'imagined' social presence could relate to external cues. Their example of vignettes used to describe imaginary situations for social psychology experiments is straightforward because it implies an external stimulus; however, in their separate example of a child's imaginary friend, it is difficult to see how one might have low awareness ("copresence"), when the imagined social presence of the friend is called into existence through thought. This example undermines their theoretical distinction between social presence as a property of the medium and copresence as a subjective sensation that others are present.

Nowak's definitions of social presence and copresence [Now01] draw directly from other theoretical discussions [BHB02, BBHS01, Bio97, Lom00]. Her concern is with finding appropriate criteria for evaluating experiences of mediated communication, the core construct of interest being the notion of connection with another mind. This construct echoes Biocca and colleagues' discussion of social presence as the access to intelligence [BHB02, BBHS01, Bio97]. Nowak proposes evaluation criteria based on some dimensions of presence discussed by Lombard [Lom00]: presence as transportation (concerning the spatial aspects of presence), copresence and social presence (concerning the social aspects of interaction). While retaining the dimensions of presence proposed by Lombard, Nowak argues that the "illusion of nonmediation" [LD97] may not be a necessary condition for social presence and copresence: "The notion that awareness of mediation prohibits the sense of presence may not apply to dimensions of presence involving communication goals" [Now01, p.3]. In short, the awareness of mediation influences, but does not prohibit, the social aspects of presence.

Nowak sees both social presence and copresence as closely related yet separate ways of evaluating mediated experience:

> "The conceptual description of these concepts appears to be the same or very similar. The indicators of social presence considered here ask about people's perceived ability of a medium to provide social presence, and do not directly measure the sense of another person, which copresence attempts to" [Now01, p.12].

This distinction between social presence as a perceived property of the medium and copresence as a subjective sense of being together closely parallels that made by Swinth and

Blascovich [SB02]. Table 2.15 summarises and compares the distinctions made between the terms. In all three cases there is an attempt to account for both the terms copresence and social presence, though with differing degrees of departure from the original definitions proposed by Short, Williams and Christie [SWC76] and Goffman [Gof63].

Table 2.15: Conceptual distinctions between copresence and social presence

	Copresence	Social Presence
Biocca & Harms 2002 [BH02]	Awareness of mediated body (subset of social presence)	Access to intelligence, Intersubjectivity
Swinth & Blascovich 2002 [SB02]	Perception & awareness that others are present (experience)	Cues signalling presence of social entity (medium)
Nowak 2001 [Now01]	Sense of another person (experience)	Perceived capacity of medium to transmit cues (medium)

2.5.8 Internal and external determinants

As in the case of presence, individuals may vary in their propensity towards social presence. Lee and Nass claim that in addition to external properties of the medium, individual differences are likely to influence social presence; these include the willingness to suspend disbelief, prior experiences with media, gender and mood [LN01]. Blascovich proposes other possible moderators of social presence [Bla02, p.136]:

1. *demographic*: race, socioeconomic status, religion;
2. *dispositional*: personality, temperament, intelligence;
3. *social*: family, group membership, social power.

Researchers investigating social responses to virtual humans can therefore benefit from collecting explanatory variables. A number of such variables were collected in the experiments described in this thesis including gender, age, occupational status, technical expertise and social anxiety (see Chapters 4 and 5).

A number of external factors relating to avatar fidelity have been investigated as possible determinants of social responses. These include agency, appearance and behaviour. Some examples of related studies and the measurement approaches taken are discussed below.

2.5.9 Measurement approaches

Relatively little has been written on the measurement of social presence and copresence in CVEs, and as mentioned by Schroeder no research survey has been published [Sch02a]. The study of social responses to avatars and agents in CVEs is still at an early stage, and there is

as yet no standardised questionnaire for social presence or copresence. Table 2.16 illustrates a range of definitions and measurement approaches taken in the subjective assessment of social presence and copresence.

Table 2.16: Summary of subjective measurement approaches for social presence

Definition	Questionnaire Items	Source
Perceived ability of medium to connect people	9 items, Likert 7-point scale (adapted from Short, Williams and Christie, 1976)	Nowak and Biocca [NB01]
Sense of connection to conversation partner	Unknown number of items, including Impersonal/personal, unsociable/sociable. Semantic differential scale (adapted from Steinfield, 1986)	Nowak and Biocca [NB01]
Perceived effectiveness of the medium in connecting people	5 items , Sliding scale (adapted from Short, Willams and Christie)	Nowak [Now01]
Sense that other intelligent beings coexist and interact with you	Likert 10-point scale	Lee and Nass [LN01]
Perceptions of the medium's capacity to transmit social cues	4 bipolar 7-point scales : Unsociable-sociable, insensitive-sensitive, impersonal-personal, cold-warm (after Short, Williams and Christie, 1976)	Sallnäs [Sal99]

Lee and Nass [LN01] present findings from a study investigating social presence responses to synthetic speech varied to reflect three different aspects of social interaction: similarity-attraction, consistency-attraction and extraversion. Their social presence index is composed of four questions concerning the mediated experience, as opposed to the medium itself. Their findings suggest that a greater sense of social presence is engendered when the synthesised voice suggested extraversion and similarity to them. Their findings are consistent with expectations from Reeves and Nass' theory of the medium as social actor [RN96].

Sallnäs describes a study investigating the effect of haptic interaction on perceived presence, social presence and task performance [Sal99]. Findings suggest that haptic force feedback had a positive but not significant effect on perceived social presence, measured using four semantic differential scales from Short, Williams and Christie [SWC76]. In related research on the role of touch, Basdogan et al. [BHSS00] use eight questionnaire items to measure copresence, expanding on the three items proposed earlier by Slater et al. [SSUS00]. The items cover a range of issues including the sense of being together, the transparency of the interface, task focus, sense of harmony, the degree of similarity to performing a joint task in physical space, and finally the sense of interacting with a person rather than a computer interface.

Slater et al. present a study on small group interaction in CVEs, where three items are

Table 2.17: Summary of subjective measurement approaches for copresence

Definition	Questionnaire Items	Source
Feeling of connection between people	26 items, Likert 5-point scale Two separate scales measuring intimacy, involvement and immediacy (adapted from Burgoon and Hale, 1987): 1. Participant's self-reported copresence (11 items) 2. Perception of partner's copresence (15 items)	Nowak and Biocca [NB01], Nowak [Now01]
Sense of being with other people in the VE	3 items, Likert 7-point scale	Slater et al. [SSUS00]
Sense of being in the same room as partner	2 items, Likert 5-point scale	Widerström et al. [WAS+00], Schroeder et al. [SSA+01]
Sense of being with partner as opposed to interacting with a computer interface	8 items, Likert 7-point scale(Partly adapted from Slater et al. [SSUS00])	Basdogan et al. [BHSS00]. Later adapted by McLaughlin et al. [MSP+03]

used to measure copresence [SSUS00]. These focus on the sense of being together in the virtual space, and the degree to which that sensation is similar to being together in a physical space. Schroeder and colleagues report on a series of studies involving a joint puzzle-solving task [WAS+00, SSA+01]. Here, the sense of copresence is equated with the sense of being in the same room or computer-generated environment, and is measured using two items that specifically concern shared space.

Although several research studies in the literature on CVEs have considered social presence and copresence, very few have investigated the impact of avatar fidelity. Nowak and Biocca present a study investigating the impact of agency and anthropomorphism on presence, copresence and social presence [NB01]. The questionnaire items are adapted from previous studies as illustrated in Table 2.17. There is some terminological confusion, since copresence and social presence are measured using separate questionnaire items, yet both are defined as access to another mind . This confusion is also reflected in the measures used. Copresence concerns the dual "feeling of connection between two people" and social presence is divided into two dimensions: the perceived capacity of the medium to transmit necessary cues, and the sense of connection to the conversation partner. It is unclear how this second dimension of social presence differs from copresence, and this blurring is inconsistent with the conceptual distinction made between the constructs by Nowak [Now01].

2.5.9.1 Summary: presence, social presence and copresence

This section has discussed presence, social presence and copresence. There are many open questions about the definition, measurement, and determinants of these constructs, as well as their possible interrelationship. One issue common to discussions of all three terms is the need to conceptually distinguish people's experiences of interaction from properties of the medium.

This section has introduced a range of approaches to the definition of social presence and copresence. In essence what all definitions aim to capture is the sense of being in the company of another person during the course of mediated interaction. Though this notion appears straightforward, there are unresolved questions about how it should be operationalised, and how it should be distinguished from its possible correlates.

This thesis will use the term copresence to describe the aggregate of individual responses leading to a sense of being with another social entity. Individual responses treated as indicators of copresence will include personal contact, partner evaluation, attributed sentience, and spatial copresence, which refers to the explicit sense of being in a shared space.

2.6 Chapter Summary

This chapter has been divided into four main sections. The first section identified the motivations for conducting the research by presenting the potential strengths of CVEs as a communications medium. Particularly when experienced immersively, CVEs can offer a multisensory, spatial experience, but cannot compare with VMC in terms of visual fidelity. The second section addressed current research in online communities and research institutions, underlining the limited expressive potential of avatars. The third section discussed the technical constraints that hinder the development of expressive avatars, calling for a need to prioritise which aspects of avatar fidelity are needed for improved communication. It also discussed related research concerning the impact of minimal fidelity on perceptions and social responses. It concluded with a summary of the approach taken in this research, namely to explore the lower boundaries of avatar fidelity with a view to maximising people's sense of being together in a shared space. The fourth and final section discussed the definition and measurement of presence, social presence and copresence, concluding with a definition of the term copresence as it will be used in this thesis.

The experiments presented in this thesis attempt to address some gaps in current research. The experiment presented in Chapter 4 investigates the impact of minimal responsiveness, which has received little explicit attention in the literature. The two experiments presented in Chapter 5 extend previous research on the impact of eye gaze in avatars by exploring the

impact of head and eye animations on perceptions of dyadic interaction. The second of these addresses the question of whether visual and behavioural realism are independent from each other, or whether their effects are intertwined.

The purpose of this chapter has been to contextualise the experiments in terms of related research on CVEs, virtual humans, presence and copresence. The following chapter on methodology will discuss the approach taken towards experimental design and analysis.

Chapter 3

Methodology

In discussions of research approaches, theorists have signaled the need to differentiate between methodology (the general approach taken to addressing a research problem) and methods (the specific techniques used to gather and analyse data) [Hen96, Wil01]. In this view, methodology is a higher-level concern that is driven by the researcher's aims and epistemological stance (underlying theoretical assumption about the basis for knowledge). The research presented in this thesis comprises three experiments investigating responses to virtual humans representing different levels of fidelity. The experimental, or hypothetico-deductive approach, is associated with a research tradition that privileges quantitative methods of data collection and analysis. However, a combination of quantitative and qualitative approaches was used to analyse participants' perceptions of their interactions with virtual humans. Some researchers view the fusion of quantitative and qualitative approaches as problematic, since they represent fundamentally divergent epistemological positions [CK98, Sci98]. Others, however, view them as complementary and mutually beneficial [Hen96, PG98, Sti93, Wil01]. The view of this author is that particularly in the case of new areas of research about which relatively little is yet known, such as avatar-mediated communication, there are benefits to partnering an experimental approach with discovery-oriented qualitative methods. The discussion about combining quantitative and qualitative approaches will be taken up in Section 3.2.3.1.

This chapter will focus on the choice of methods used to address the research questions. The experiments focused on distinct aspects of fidelity, but shared many similarities in terms of data collection and analysis. Section 3.1 will focus on methods of data gathering. It will detail the strategies used to design and pilot the experiments, as well as the experimental procedures common to all three. Section 3.2 will focus on methods of data analysis. The data was in three forms: subjective questionnaire data, subjective interview data and objective psychophysiological data (heart rate and EDA). Methods of analysis for each type of data will be described in turn. The chapter will conclude with a summary table of the types of data collected in each

experiment, and the corresponding methods of analysis used.

3.1 Data Collection

This section will cover aspects of experimental design and procedure common to all three experiments. Chapters 4 and 5 will describe additional details particular to each experiment.

3.1.1 Defining the research goals and expectations

All three experiments had a common theme, namely the impact of avatar fidelity on social interaction. However, different aspects of fidelity as well as different responses were explored in each experiment. The general expectation was that the greater the level of fidelity, the more the virtual humans would be seen to contribute to the experience and the more they would elicit social responses from participants. However, one challenge in this emerging area of research is that, just as there exist many questions about the impact of virtual humans, so are there open questions about what constitutes a social response. The first step in designing the experiments was therefore to define the specific research question in terms of the exact independent and dependent variables of interest.

3.1.2 Defining the independent and dependent variables

The broad purpose of the research was to investigate the impact of avatar fidelity on a selection of responses. For each experiment a specific aspect of fidelity was chosen for investigation, for instance, eye gaze realism. This independent variable was then manipulated to create different conditions for comparison, for example less realistic and more realistic eye gaze. The dependent (or response) variable was operationalised by defining how the responses of interest were to be measured. It was divided into different indicators, depending on the focus of the experiment. For instance, rather than simply investigating 'social presence', a number of constructs that were believed to be indicators of social presence were selected, for example, the degree to which the virtual humans were responded to as people rather than as a computer interface.

The central focus of the thesis was on subjective responses, therefore the data was gathered primarily through questionnaires. Each indicator consisted of a number of questionnaire items. Some of these items were based on previously published research, while others were developed during the course of the research through the process of piloting questionnaires and analysing interview data from previous experiments.

3.1.3 Experimental design

All three experiments used a between-groups factorial design. Each *factor* in the experiment represented a major independent variable under investigation, namely a specific aspect of avatar

fidelity. This could in turn be subdivided into different *levels*. For example, the experiment presented in Chapter 4 investigated a single factor, responsiveness, with four different conditions representing increasing levels of responsiveness to the human participant, as in Table 3.1. A later experiment (presented in Section 5.2) used a two-by-two design, as in Table 3.2.

Table 3.1: One-factor design with four levels

	Factor 1
Level 1	Condition 1
Level 2	Condition 2
Level 3	Condition 3
Level 4	Condition 4

Table 3.2: Two-by-two factorial design, each factor having two levels

		Factor 1	
		Level 1	**Level 2**
Factor 2	**Level 1**	Condition 1	Condition 2
	Level 2	Condition 3	Condition 4

Between-groups experiments have the disadvantage of requiring twice the number of participants as compared to within-groups experiments. This added to the cost and logistical complication of recruiting participants, particularly as two of the experiments required pairs of gender-matched participants who did not know each other prior to the experiment. Another disadvantage of between-groups designs is that it is not possible to completely control for participant variables. However, an effort was made to remedy this by randomly allocating participants to each condition, and by collecting as much explanatory data as possible.

On the other hand, there are also advantages to using between-groups designs. Order effects are not a concern and it is possible to use the same task and materials for each group. Another key advantage is that participants only experience one experimental condition. They therefore have no means of comparing between conditions and biasing their responses based on their assumptions about what the experiment is about (demand characteristics).

In this research, employing a between-groups design also offered practical advantages in terms of timing and scheduling. Within-groups experiments require a pause between conditions to minimise participant boredom and fatigue. This pause is of even greater importance in the case of experiments using immersive equipment (such as the Cave and HMD) where it is necessary to keep session times brief in order to minimise the risks of simulator sickness. Given that

the laboratory equipment used was expensive and in high demand, it was essential to run the experiments in the shortest possible time, scheduling four to six back-to-back sessions per day whenever feasible. Eliminating the pauses between sessions that would have been necessary for a within-groups design therefore meant saving precious time. For all the above reasons it was decided that the advantages of a between-groups design outweighed the disadvantages of requiring a larger sample size.

3.1.4 Piloting

The process of piloting was an essential step in the iterative process of designing each experiment. A small sample of people were invited to participate in the pilot sessions. The purpose of these sessions was to evaluate different aspects of experimental design including procedure, task, questionnaire items and interview questions. Numerous modifications were made to the experiment as a result of observations and feedback from these sessions. Naturally, no data from the pilot sessions was included in the results.

In addition to improving various aspects of experimental design, piloting also served as a dress rehearsal, helping to make the experimental procedure more familiar to the experimenters. This process of familiarisation was of paramount importance in ensuring that a standardised procedure was maintained throughout each experiment.

3.1.5 Ethical considerations

Prior to running experiments at UCL it was necessary to submit details of the experimental design and procedure for clearance from the ethics committee. This was done towards the end of the piloting phase, when most details of the experiment had been finalised. Guidelines for ethical conduct in research apply equally to quantitative and qualitative approaches [CK98, Wil01] and include the following points:

1. *informed consent*: Participants must be given sufficient information about the research procedures so that they can make an informed decision about participating. They must also give their consent to participate prior to the data collection phase;

2. *the right to withdraw*: participants must be made aware that they are free to withdraw at any time;

3. *confidentiality*: any information acquired about the participants during the course of the research must be kept confidential;

4. *deception*: participants must not be deceived;

5. *debriefing*: after data collection, participants must be informed about the aims of the research.

All three experiments complied with the above guidelines.

3.1.6 Experimental procedure

Upon arrival, participants were greeted in a reception area by two experimenters (the author and a colleague). In cases where two people were participating at the same time, one experimenter was assigned to 'mind' each participant for the duration of the session.

Participants were given an instruction sheet to read that detailed the experimental procedures and, where relevant, the possible risks associated with using immersive equipment. They were then asked to sign a consent form, in accordance with the procedures required by the ethics committee. Where relevant, the form explicitly asked for written consent to being audio and videotaped. Participants were informed that all data, including audiovisual records, would be confidential and would only be used for the purpose of data analysis. They were also instructed that they were free to withdraw from the experiment at any time and without giving a reason for withdrawing.

Participants were then given time to prepare for their task. Depending on the focus of the experiment, this included either preparing a scenario for conversation, or performing a navigation training task in an immersive environment, or both. Once they felt ready to proceed they were reminded of the amount of time they would have to perform their experimental task, and that at the end of the task the experimenter would return to guide them through the next stage.

During the task, the experimenters quietly observed participants. In the interests of a standardised procedure, participants were stopped at the end of the assigned time period regardless of whether the task had been completed. This was of particular importance in the experiments involving immersive equipment, again to minimise the risk of simulator sickness.

After completing the task, participants filled out a post-questionnaire about their experience. Sessions concluded with a semi-structured interview. Finally, participants were debriefed as to the purposes of the research, and were asked not to speak about their experience for the planned duration of the experiment to avoid revealing its objectives to other potential participants.

This section has detailed the aspects of procedure that were common to all three experiments. However, given the different aims of each experiment, there were necessarily a number of variations in procedure. Some of these are outlined in a Table 3.8 at the end of this chapter. Further details are given in the descriptions of the individual experiments in Chapters 4 and 5.

3.1.7 Semi-structured interviews

Semi-structured interviews were conducted individually with participants at the end of each session. Semi-structured interviews differ from structured interviews in key ways [Smi95]. Structured interviews, such as those used in the collection of census data, have much in common with questionnaires in that they cover a set of predetermined questions with pre-coded response categories. Questions are typically short and are often read out loud to the respondent in a specific order. While this presents advantages in terms of control, reliability and speed, the drawback is that little or no room is left for exploring avenues that do not fit into the predetermined categories. Interviews can also appear formal to respondents.

In contrast, while formal in terms of their research program aims and their assignment of roles to the interviewer and interviewee, semi-structured interviews preserve some of the features of informal conversation. Particular emphasis is placed on establishing rapport with the interviewee [Smi95, Wil01]. This sense of rapport implies ethical responsibilities, and interviewers need to be sensitive to respondents' willingness to talk about given topics, avoiding those that appear awkward or uncomfortable. Semi-structured interviews are designed to generate a detailed description of respondents' perceptions of a given issue. They are therefore flexible, and it is not necessary for questions to be asked in a set order or even to be phrased in the same way with each participant. This flexibility enables respondents to focus on those topics of greatest interest and relevance to them, yielding richer and more varied data. In this sense, participants play a central role in shaping the direction of the interview.

It is the responsibility of the researcher, however, to ensure that the interview does not stray from the general research question. A carefully constructed interview agenda (or schedule) can help in this [Smi95, Wil01]. Interview agendas are designed in advance to identify logically ordered themes. While some interviewers prepare a selection of topics, others prefer to word exact questions in advance. It is also advisable to design 'open' questions that cannot be satisfied by a 'yes/no' answer, and to avoid using jargon or asking leading questions.

3.1.7.1 Taping and transcription of semi-structured interviews

Although taping interviews can potentially make some respondents less comfortable, it has the advantage of preserving a more complete record of the interview than would be possible when taking notes [Smi95, Wil01]. Semi-structured interviews are often used in conjunction with qualitative methods of analysis to open new theoretical avenues [Smi95]. Depending on the type of qualitative analysis to be conducted, transcription can follow precise notation, signalling such details as pauses, false starts, interruptions and intonation. The method of qualitative analysis used in this research was grounded theory [SC98]. Grounded theory is concerned with

the content of what is being said and it is therefore only necessary to transcribe the words themselves [Wil01]. The interviews used for the grounded theory analysis were transcribed by the author. This was done with a view to maximising familiarity with the raw data.

3.2 Data Analysis

This section covers the methods used to analyse the data gathered in the three experiments. As mentioned earlier, the data included subjective questionnaire and interview responses as well as objective data in the form of heart rate and EDA. The first section covers the statistical methods used to analyse questionnaire data, and applies to all three experiments. The second section focuses on the analysis of psychophysiological data and applies to the experiment on responsiveness detailed in Chapter 4. The third and final section on the grounded theory analysis of interview data applies to the first experiment on eye gaze, covered in Section 5.1.

3.2.1 Quantitative analysis of questionnaire data

The questionnaires used Likert-type scales. Likert scales are of course ordinal and it is not appropriate to analyse them as interval responses. The method was used as in [SSMM98] which provides a conservative analysis of the responses based only on count data. This has the advantage of never using the dependent variable ordinal questionnaire responses as if they were on an interval scale.

Each response variable is constructed from a set of n questions. 'High' responses are defined by a cut-off value on the Likert scale; for example, on a scale of 1-7, values of 6 or 7 are classified as a high (or positive) response. The number of positive responses (r) out of the n possible positive responses are then counted for each set of questions. Therefore each response variable is a count out of n possible high scores.

Under the null hypothesis of randomly and independently assigned responses, r has a binomial distribution and therefore logistic regression can be used for the analysis as to how r varies across the main condition and with respect to other variables [MN83]. In the case where the right-hand-side of the regression consists of only one factor (for example, the condition) this is equivalent to a one-way ANOVA but using the more appropriate Binomial distribution rather than the Normal. The *deviance* is the appropriate goodness of fit measure for this regression model, and has an approximate chi-squared distribution with degrees of freedom depending on the number of fitted parameters. A rule-of-thumb is that if the deviance is less than twice the degrees of freedom then the model is a good fit (at the 5% significance level). The change in deviance as variables are deleted from or added to the current model is especially useful, since this indicates the significance of that variable in the fitted model. Here a large change of

deviance indicates the degree of significance, i.e., the contribution of the variable to the overall fit. The analysis was carried out using the GLIM system [Gen03].

3.2.2 Quantitative analysis of psychophysiological data

Although the primary focus of this research is on subjective perceptions of avatars, the possibility of using objective measures to study responses to virtual humans was also explored. Psychophysiological data was collected in the experiment on responsiveness (Chapter 4) in order to investigate whether additional understanding about responses to virtual humans could be gained by triangulating subjective and objective methods. Details on the specific apparatus and procedures for data gathering can be found in Section 4.2.3.

Research into physiological measurement of human responses has been driven by an interest in the interconnection between mind and body [And80, CT90]. Where *physiological psychology* is concerned with the impact of physiology on psychological processes, *psychophysiology* is concerned with the impact of psychological processes on physiological responses. The premise is that there is a connection between mental and physical states, and that reactions to certain experiences can therefore be expressed in measurable physiological changes. More explicitly, the premise is that state of mind can influence the body directly [And80]. It has been argued that in there is no unifying theory in the field of psychophysiology, rather a collection of concepts that often contradict each other [And80]. Research using psychophysiological methods to investigate human-computer interaction is still in an exploratory stage. Nonetheless, there has been some interest in presence-related research in investigating the use of psychophysiological measures as objective indicators of presence in VEs [DKF02, Mee01, MIWB01].

Heart rate and electrodermal activity (EDA) were used by Dillon, Keogh and Freeman to investigate responses to different types of visual displays [DKF02]. Meehan et al. used heart rate, EDA and skin temperature to investigate responses to a virtual 'pit' room containing a 20-foot drop to a room below [MIWB01]. The premise was that to the degree that the virtual room could evoke presence, it would trigger similar psychophysiological responses to a comparable real-world stimulus. In this case, results confirmed expectations that a fear of heights would result in an associated increase in heart rate and EDA, and a decrease in skin temperature. Note however that participants were deliberately exposed to an extreme stimulus designed to evoke a high level of anxiety or arousal.

It is arguable that psychophysiological changes may not be equally pronounced in the social scenarios of interest in this thesis. Nonetheless, the same 'arousal' responses were of interest in this research. A range of psychophysiological responses can be measured, including pupil dilation, heart rate, EDA and skin temperature. The two responses used in this research

were heart rate and EDA.

Heart rate (HR) can be measured in a variety of ways, most commonly via blood volume pulse (BVP) and electrocardiogram (EKG). When immersed in a Cave, participants must be free to move physically around the space. For this reason the sensors used by Wilson and Sasse in recent experiments on the impact of media quality on user cost were not appropriate, since these required the participant to sit still for an accurate reading [WS00]. Responses were therefore measured using the procedures reported in [MIWB01], since these were tailored to a comparable situation involving a mobile participant exploring a VE. A three-electrode electrocardiograph was attached to the chest, as this had the advantage of reducing motion artifact. Heart contractions are accompanied by electrical changes recordable from the surface of the skin, so that EKG readings can be converted into heart beats.

Electrodermal activity (EDA) has also been referred to as galvanic skin response (GSR). It is produced by two types of sweat glands: the apocrine glands (located in the armpits and genital areas, which respond primarily to temperature) and the eccrine glands (located all over the body and found in particularly high density on the palms of hands and soles of feet) [CT90]. The most commonly used measure of EDA is skin conductance, because it simplifies the interpretation of results; conductance is positively correlated with the level of secretion of the sweat glands, and therefore linearly illustrates a corresponding increase in arousal. Skin conductance was measured by passing a weak electrical current across the skin using bipolar electrodes placed on adjacent fingers [MIWB01].

It is common for researchers to report their psychophysiological findings in terms of overall means [CT90]. In Meehan's experiments [MIWB01], the impact of the pit room was calculated in terms of the increase in mean level of HR and EDA from the neutral 'training' room to the experimental 'pit room.' The same method of analysis was applied here, since there was a direct parallel in terms of experimental procedure. For each participant, mean levels of HR and EDA in the training room were compared to mean levels in the richer experimental environment containing avatars. Results were analysed in terms of an increase in mean levels.

3.2.3 Qualitative analysis of interview data using grounded theory

This section focuses on the qualitative analysis of interview transcripts from the first experiment on eye gaze presented in Section 5.1. The rationale for using a qualitative method is that while questionnaire data allows for the compact analysis of a set of predetermined variables, the scope of what can be learned is limited by the fact that participants' responses are restricted to a set of preconceived themes. In a new research area such as avatar-mediated communication, it is conceivable that the researcher might choose to focus on issues that are only of

peripheral concern to respondents, and ultimately to end-users. Some studies on interaction in shared virtual environments have aimed to address this concern by including free response items in questionnaires, or by conducting interviews with participants. However, these have usually been reported in anecdotal support of quantitative findings or in suggesting avenues for future research, without conducting a structured analysis of the qualitative data (see for example [SHS+00]). The rationale for conducting a qualitative analysis in this research was to focus the investigation in light of participants' own perspectives.

The grounded theory method seemed particularly appropriate because it is tailored to the exploration of complex phenomena, as well as phenomena about which relatively little is yet known. It is also designed specifically to analyse those aspects of a topic deemed important and relevant by respondents [SC98]. One point to emphasise is that analysis is painstaking and time-consuming, and time constraints meant that it was impractical to apply it to the interview data from all three experiments. Since grounded theory is designed to be a discovery-oriented, theory-generating method, it was applied to the data from the first experiment with a view to clarifying research questions for the successive studies.

Grounded theory has been referred to as both a *method* and a *theory* [Wil01]. For this reason a description of its procedures cannot be entirely separated from methodological and epistemological concerns. The following section will contextualise grounded theory within a wider debate about qualitative methods. A description of how the procedures were adapted for the purposes of this analysis will follow.

3.2.3.1 Theoretical background: contextualising grounded theory within the quality/quantity debate

In discussing qualitative research, several theorists argue that the choice of research methods is inextricably linked with epistemological views [CK98, Sci98]. Since the mid-twentieth century, the human sciences have witnessed lively debates on the relative merits of quantitative and qualitative research. The two approaches are often seen as representing fundamentally opposed views about the theoretical basis of knowledge. Quantitative research is traditionally associated with a positivist, realist view of the relationship between the knower and the known; it privileges objective accounts of a reality that is believed to be external to the researcher. Conversely, qualitative research is associated with a social constructionist (or constructivist) view, proposing that knowledge is not simply discovered, but actively constructed by researchers' interactions with data [Hen96, Pid96, Wil01].

Table 3.3 illustrates that the differences between quantitative and qualitative approaches span from underlying theoretical positions to practical concerns about how to gather and anal-

Table 3.3: Comparison between quantitative and qualitative approaches

	Quantitative	Qualitative
Epistemological association	Positivism, Realism, Empiricism	Relativism, Constructionism (Constructivism)
Emphasis on	Prediction	Description and interpretation
Knowledge is generated	Hypothetico-deductively	Inductively
Point of departure	Hypotheses	Data (to be analysed as far as possible without imposing expectations or preconceptions)
Aim	Hypothesis testing (Received-view)	Hypothesis generation (Discovery-oriented)
Form of data analysed	Numerical, e.g. questionnaire data	Primarily textual, e.g. interviews and narratives
Methods of data gathering	Experiments (primarily laboratory settings)	Semi-structured interviews, participant observation, diaries, focus groups, case studies (primarily naturalistic settings)
Examples of methods of data analysis	Statistical analysis of questionnaire results	Grounded theory, content analysis, discourse analysis, protocol analysis, ethnography

yse data. There are at least two levels to the qualitative/quantitative debate. These have been referred to as the 'maximalist' (epistemological) debate on the one hand, and the 'minimalist' (technical) debates regarding practical choice of research methods, on the other [Hen96, p.29].

If the quantitative/qualitative debate is viewed on the epistemological level alone, then it is not clear how the two approaches could ever be reconciled, since they are rooted in fundamentally divergent and incompatible theoretical positions. Some theorists have privileged the epistemological debate over the technical, arguing that the choice between quantitative and qualitative approaches is too often decided on the pragmatic level alone, ignoring the philosophical commitments associated with each approach [Sci98, p.37].

However, not all researchers agree that viewing the quality/quantity debate as a clash of paradigms is necessarily productive [Hen96, Pid96, Sti93]. The two approaches can be seen as complementary and there is room for a fruitful symbiosis, provided they are judiciously combined:

> "Accepting the value of the distinction between quantitative and qualitative paradigms should not be taken to imply that quantitative and qualitative perspectives and methods are mutually exclusive, or that there is a one-to-one relationship between 'quantity' and epistemological realism, on the one hand, and 'quality' and

> constructionism, on the other. Researchers often wish to argue for a principled mixture of quantitative and qualitative methods" [Hen96, p.30].

There are different ways in which the approaches can be reconciled. On the pragmatic level, even within the positivist view there is room for employing qualitative methods for early, explorative studies such as pilot experiments [Sti93]. Nonetheless, to take this position is to sidestep the epistemological debate and to ultimately deny that qualitative methods have value in their own right. A more satisfactory solution in theoretical terms is proposed by Charmaz [Cha95], who presents grounded theory as a way of bridging positivist assumptions with interpretative analyses. In contrast with the positivist approach that claims to study 'external' reality, interpretative analyses aim to understand and explain personal views of lived experience from the 'inside'. Charmaz argues that grounded theory investigates these personal experiences while still remaining empirical:

> "Grounded theorists typically provide dispassionate, objectivist accounts of their data and assume that by being objective observers they will discover processes in an external world of their research participants that remains separate from themselves" [Cha95, p.31].

It is therefore argued that grounded theory can be used by researchers who subscribe to positivist as well as constructionist assumptions. Willig agrees, cautioning however that different epistemological positions call for different emphases in quality control and the process of analysing and reporting data [Wil01]. This issue of quality control in grounded theory will be taken up again in Section 3.2.3.5.

To focus exclusively on the quality/quantity debate is to lose sight of the range of methods within qualitative research. Although these vary in terms of aims and procedures, they share a common concern with the meaning people attribute to their experiences [Sti93, Wil01]. They also privilege the process and quality of experience over prediction and identifying cause-effect relationships. Another common feature is the goal of broadening investigation beyond a set of predefined variables. In summary, the main argument for employing qualitative approaches is that human experience has multiple dimensions and cannot be adequately represented by reducing it to a range of scales [Sti93].

3.2.3.2 Grounded theory: guiding principles

Barney Glaser and Anselm Strauss first introduced grounded theory in 1963. They argued for the adoption of an inductive, theory-generating qualitative approach against a prevailing climate of quantitative research in sociology. Their concern was that the hypothetico-deductive method,

with its emphasis on testing hypotheses through empirical studies, did not leave room for the development of new theories through research. Against a top-down approach in which theories were tested against data, they argued instead for a bottom-up approach in which theory should emerge directly *from* data.

Grounded theory shares several tenets with the qualitative paradigm as a whole, most significantly in its commitment to discovery [Pid96, Sti93, Wil01], as well as its emphasis on grounding the investigation in participants' own accounts of their experiences [Pid96]. Grounded theory is a nonmathematical method of analysis that can be applied to a range of textual as well as visual data gathered by a variety of means. Textual data can include field-work observations, case-study documentation, archival research and, perhaps most commonly, semi-structured interviews [Pid96].

Among the qualitative methods available for analysing textual data, grounded theory most commonly risks being confused with content analysis, although the two methods have contrasting aims and procedures. Content analysis involves counting incidences of predefined elements in the data. As such, it is more closely linked with the hypothesis-testing aims of quantitative analysis, and accordingly reflects a concern with validity and inter-rater reliability. Conversely, grounded theory aims to generate hypotheses and theories from the data itself. Its primary concern is not with counting occurrences, rather with investigating interconnections. Another key point of divergence from content analysis specifically, and from quantitative methods more generally, is the intermingling of data gathering and analysis. Rather than representing distinct phases in the research, they contribute to an iterative process where ongoing analysis shapes subsequent sampling decisions. Some of the key ways in which grounded theory departs from the hypothetico-deductive approach are summarised in Table 3.4.

Table 3.4: Key differences between the hypothetico-deductive and grounded theory approach

Hypothetico-deductive approach	**Grounded theory approach**
Theory is first formulated, then tested against data	Theory emerges from the data
Aim is hypothesis testing	Aim is discovery and theory generation
Random sampling	Sampling decisions are driven by the emerging analysis
Clear separation between data gathering and analysis stages of research	Data gathering and analysis are intertwined, forming an iterative process of that ends at the point of theoretical saturation

There is, however, a confusing proliferation of methods within grounded theory [Dey99, Wil01]. This can partly be explained by the theoretical disagreement between its original pro-

ponents, Glaser and Strauss. Strauss, with student Corbin, developed the method in new directions that in Glaser's eyes broke with the original premises of grounded theory [Dey99, Wil01]. Where Strauss and Corbin stressed the importance of constantly verifying the emerging theory against the data itself, Glaser held that the concern of grounded theory was purely to generate hypotheses, not to verify them. Strauss and Corbin also introduced a new element in the method called the 'paradigm,' a conceptual scheme designed to link the twin themes of structure and process in analysis.

The analysis conducted in this research was based primarily on the procedures set out in Strauss and Corbin's introductory text [SC98], considered by many "the standard introduction to grounded theory" [Dey99, p.13]. As mentioned above, there is a wide range of interpretations on how to conduct grounded theory, and additional texts helped to elucidate the set of procedures for data collection and analysis [Cha95, Dey99, Hen96, Pid96, PH96, Sti93, Wil01]. The following section will introduce the procedures and tools used in grounded theory analysis. Next there will be a discussion of approaches to quality control. An explanation of how the analysis departed from standard procedures will follow.

3.2.3.3 Grounded theory procedures: the three stages of coding

Strauss and Corbin describe grounded theory as a process of conceptualisation and abstraction that involves breaking the raw data down and putting it back together in new ways [SC98]. The analytical process consists of three phases: *open coding, axial coding* and *selective coding*.

Table 3.5: Summary of main analytical tasks associated with each of the three phases of coding in grounded theory

	Open coding	**Axial coding**	**Selective coding**
Emphasis	Classification	Identifying relationships	Refining the theory
Main tasks	*Labeling*: Naming concepts *Microanalysis*: Defining and developing categories in terms of their properties and dimensions.	Systematically linking categories with their subcategories, and with other categories at the dimensional level. Coding for structure and process using the paradigm.	Identifying the core category and storyline.

Although the three phases are logically distinct and serve different purposes, they often occur concurrently, most typically in the simultaneous use of open and axial coding. Table 3.5 summarises the main emphases and tasks for each phase. A more detailed description of each phase of coding is given below.

Open coding: In the first stage of analysis the raw data is 'opened up' to expose the distinct ideas it contains. The first priority in open coding is to break the data down into discrete parts by identifying the basic units of analysis, referred to as *concepts*. Identifying concepts involves the detailed process of labelling the data. The granularity of labelling can vary from word-by-word coding, to coding by sentence, and finally to coding by paragraph. The analysis presented in Chapter 6 of this thesis began with a detailed labelling of all individual words that appeared to indicate significant concepts. In cases where the interview itself contained a particularly evocative word or phrase, this was directly used as an 'in vivo' code. Typically this very detailed, word-by-word coding is appropriate for the early part of the analysis, to avoid imposing preconceived classifications on the data. Later, coding can be conducted by sentence or even by paragraph.

Labelling is a form of description, and as such does not create new meaning. The process of labelling is therefore accompanied by *microanalysis*, which involves the close examination and comparison of concepts. Similar or logically related concepts are grouped into more abstract units of analysis called *categories*. Microanalysis is aided by the use of theoretical memos that record thoughts and impressions about how the emerging categories should be defined and interrelated. Through this process of microanalysis, categories are explored for their *properties* and *dimensions*, defined by Strauss and Corbin as follows: "Whereas properties are the general or specific characteristics or attributes of a category, dimensions represent the location of a property along a continuum or range" [SC98, p.117]. In other words, a dimension represents the measured extent of a property. The process of identifying properties and dimensions is particularly important since it is at the dimensional level that the connections between categories are specified.

The fundamental purpose of the open coding stage is therefore to move from description to conceptualisation:

> "In doing our analyses, we conceptualise and classify events, acts, and outcomes. The categories that emerge, along with their relationships, are the foundations for our developing theory. This abstracting, reducing and relating is what makes the difference between theoretical and descriptive coding (or theory building and doing description). Doing line-by-line coding through which categories, their properties, and relationships emerge automatically takes us beyond description and puts us into a conceptual mode of analysis" [SC98, p.66].

Moving from raw data to concepts and then to categories represents the first stage of the

process of abstraction that is central to grounded theory analysis.

Axial coding: In grounded theory, a category stands for a phenomenon. A phenomenon is a problem, issue or event that respondents deem significant in describing their experiences. The purpose of axial coding is to form more precise explanations of phenomena. During the open coding stage, various categories are identified along with their properties and dimensions. However, during this initial phase of classification it is not yet clear how they relate to each other. The central task in axial coding is to refine definitions of categories while defining their interrelationships by linking them at the dimensional level.

Strauss and Corbin emphasise that in exploring the relationships between categories it is essential to relate *structure* with *process*. Structure relates to the circumstances that explain *why* events occur, whereas process relates to *how* those circumstances are responded to (see Table 3.6).

Table 3.6: Comparing the role of structure and process in Strauss and Corbin's paradigm

Structure	Process
Circumstances leading to events *(Why)*	*How* circumstances are responded to
Conditions	Actions/Interactions

Dealing only with the *why* or the *how* of events would yield only a partial explanation of a phenomenon. Strauss and Corbin explain that "one must study both structure and process to capture the dynamic and evolving nature of events" [SC98, p.127]. In order to facilitate the analysis relating structure and process, they propose the use of what they call the *paradigm*, a conceptual scheme that helps to tease out the often subtle connections between categories. The paradigm consists of *conditions, actions/interactions* and *consequences*.

Conditions: Conditions describe the circumstances leading to an event. They can be either *micro* (closely tied to the experience in question), or *macro* (more generalised). They are classified into three types: causal, intervening and contextual. The relationship between the three is suggested by their names: causal conditions represent the events that directly influence a phenomenon, and intervening conditions alter or mitigate their impact (such as contingencies or unexpected events). Contextual conditions result from the intersection of causal and intervening conditions (Figure 3.1). Strauss and Corbin stress that this classification is useful for sensitising analysts to the complex interweaving of events and situations leading to phenomena.

Actions/interactions and consequences: Actions/interactions are the strategic responses made by individuals to conditions. They describe the way individuals handle the situations they find themselves in, and therefore relate to process. Finally, *consequences* represent the outcome

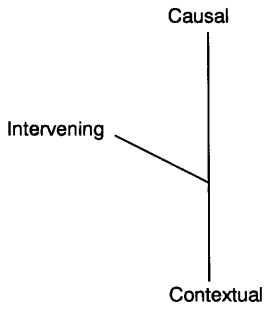


Figure 3.1: Conditions set the stage for events. Context is shaped directly by causal conditions, mitigated by intervening conditions.

of actions/interactions in response to conditions. Consequences can in turn create or alter new conditions for action/interaction, creating a cyclical process, as illustrated in Figure 3.2.

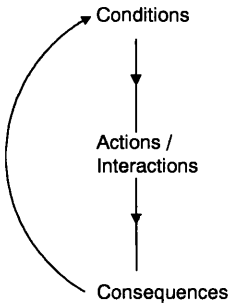


Figure 3.2: Components of the paradigm interrelate to link structure with process. Conditions set the stage for actions/interactions, resulting in consequences that can in turn shape new conditions.

By emphasising the way consequences can feed back into the system to modify conditions, the paradigm makes it possible to account for complex interrelationships between categories over time. Stiles argues that this opens new avenues for analysis:

> "Most traditional research hypotheses in psychology are stated in linear terms–one thing causes another thing. At best, hypotheses are framed in terms of interactions of two or three variables–for example, that a particular (linear) relationship holds only under stated conditions. But such lawlike generalisations may be untrustworthy if the system is non-linear, as most experiential and behavioural systems obviously are; that is, the variables interact with each other across time" [Sti93, p.197].

The connections between structure (conditions) and process (action/interaction) are not

usually explicitly stated in respondents' descriptions, and it is the responsibility of the researcher to abstract from the data in order to uncover them. Some useful indicators of the connections between the components of the paradigm are those words or phrases that indicate causation or outcome, such as: *For that reason, because, due to, when, since, what I did was, what happened was, my reaction was to.*

Selective coding: During this third and final phase of analysis, the key categories defined during axial coding are integrated and refined to form a theory. One of the key steps involves identifying the *core category*. This is the most central, abstract and descriptive category, in that it is linked to the majority of categories in the scheme and is abstract enough to contain and explain them: "A central category has analytic power. What gives it that power is its ability to pull the other categories together to form an explanatory whole. Also, a central category should be able to account for considerable variation between categories" [SC98, p.146]. There are several key requirements for the core category. It must be central (in other words it must relate to all other major categories), there must be frequent indicators pointing to it in the data, and it must be able to explain both the central point made in the data as well as the variation within it. Once the core category is identified, it then becomes possible to write a *storyline* describing the key elements and interrelationships that form the theoretical whole.

Grounded theory analysis does not aim to unearth simple cause and effect relationships between predetermined variables. Rather it aims to build a picture of how they connect with each other, In the words of Strauss and Corbin, "Identifying, sifting through, and sorting through all of the possible factors showing the nature of the relationships does not result in a simple 'if... then' statement. The result is much more likely to be a discussion that takes readers along a complex path of interrelationships, each in its own patterned way, that explains what is going on" [SC98, p.130]. The purpose of analysis is to uncover a theoretical model of how phenomena interrelate with each other.

3.2.3.4 Tools used in grounded theory analysis

Researchers are encouraged to keep detailed memos describing the emerging analysis. In addition to these, *mini-frameworks* and *conceptual diagrams* are useful visual tools for representing the connections between categories. Mini-frameworks record the intersection of two categories at the dimensional level (Figure 3.3). Conceptual diagrams are used to record emerging relationships between two or more categories and their subcategories (Figure 3.4).

In addition to these conceptual tools, software tools can also facilitate analysis. The ATLAS.ti software program [ATL03] was used for the open coding. It enables the user to import primary documents from the data pool into what is called a 'hermeneutic unit.' When selections

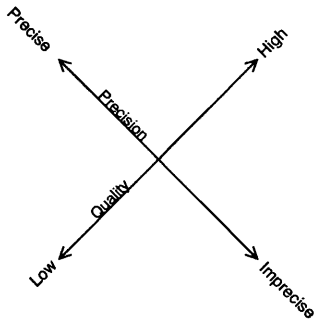


Figure 3.3: Example of a mini-framework illustrating the intersection of two categories

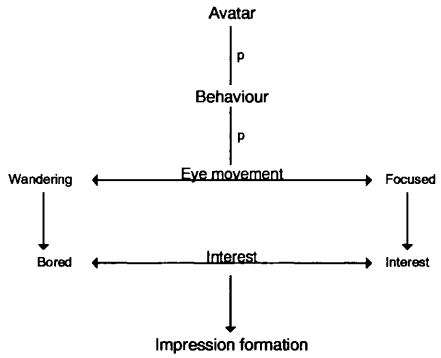


Figure 3.4: Conceptual diagram illustrating emerging relationships between a selection of categories and their properties

of text are coded, the codenames appear in the margin, as illustrated in Figure 3.5. The software is particularly useful for the open coding stage. It allows the user to code text passages, trace all occurrences of a particular code and cross-reference them with memos and other codes. Memos make it possible to keep a detailed running log of the emerging analysis linked with relevant passages in the primary documents. This greatly facilitates the 'constant comparison' Strauss and Corbin stress is essential to the process of grounded theory.

3.2.3.5 Quality control in grounded theory

This section concerns the criteria for quality control and evaluation applicable to qualitative research in general, and to grounded theory in particular. Both quantitative and qualitative approaches prize objectivity and detachment, while taking different views regarding the degree to which it is attainable in practice [Hen96, Pid96]. Qualitative research more openly acknowledges that a certain degree of subjectivity is implicit in all research:

> "Fortunately over the years researchers have learned that a state of complete objectivity is impossible and that in every piece of research–quantitative or qualitative–

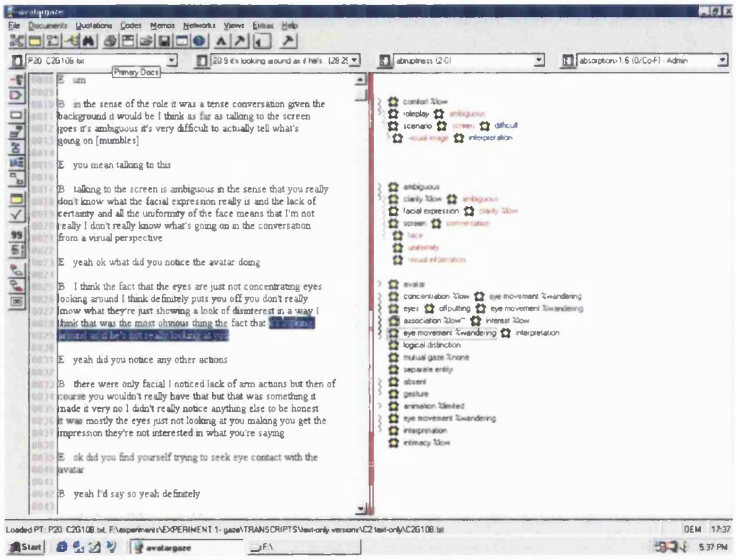


Figure 3.5: Screen shot of an ATLAS.ti session. Concept labels appear on the right of the primary document. Primary documents, codes, memos and diagrams are all searchable, as are the interconnections between them.

> there is an element of subjectivity. What is important is to recognise that subjectivity is an issue and that researchers should take appropriate measures to minimise its intrusion into their analyses" [SC98, p.43].

There is lively debate as to exactly what these measures should be [Wil01]. The fact that qualitative researchers acknowledge the subjective and creative element in research often means that the onus is on them to demonstrate the scientific validity of their findings. One of the challenges in addressing issues of quality control is that the criteria for assessment applied to quantitative analysis are not necessarily relevant or applicable to qualitative approaches [Mer98, Wil01].

Approaches to quality control in quantitative research include reliability and validity. *Reliability*, as its name suggests, considers whether the reported results can be relied upon. It concerns replicability (whether the research could be conducted elsewhere and still produce comparable results) and inter-rater reliability (whether different investigators score or interpret data in consistent or comparable ways). *Validity* considers whether the research is actually measuring what it proposes to measure. It includes both *internal* validity (whether the methods of data gathering and analysis actually pinpoint the factor under investigation) and *external* validity (whether the results can be generalised beyond the specific research context to other circumstances).

These standards are not meaningfully applicable given the premises of qualitative research. The concept of replication does not apply because of the emphasis on context in qualitative research: "The concept of replication changes meaning when the experiences and behaviour of individuals are considered as dependent upon context. No two things are exactly alike, so the similarity of any two events is an abstraction" [Sti93, p.596]. Similarly the notion of validity, in the sense of measuring objective facts, is not reconcilable with the epistemological foundations of qualitative research. Validity in qualitative research takes on a different slant, placing emphasis not on the discovery of external facts, but on the balanced understanding and interpretation of respondents' meanings [Mer98].

Several alternatives applicable to grounded theory have been proposed. *Respondent validation* (also referred to as *testimonial validation* in [Sti93]) involves ensuring that the researcher's analysis seems logical and convincing to the respondents themselves. While theoretically desirable, it is not always possible in practice to contact respondents after the data gathering stage [Pid96]. Another approach that links with inter-rater reliability is to make an 'audit trail' available to other researchers, so that the process of theory generation from raw data can be validated by others [Pid96]. A third option is *negative case analysis*; this involves actively seeking out cases that challenge the emerging theory in order to generate new insight. However, this presumes that the sampling and analysis stages are intertwined.

Some additional approaches are proposed in [PH96]. *Goodness of fit* involves ensuring that the categories fit the data through careful and exhaustive *documentation* of why phenomena are categorised in particular ways. Another issue that receives particular attention in discussions of evaluation criteria is *reflexivity* [Mer98, Sti93, Wil01]. This calls for researchers to acknowledge their role in the research by documenting how their views changed in the course of the analysis. This implies an open and 'permeable' relationship with the data, an attitude also referred to as *permeability* [Sti93]. All research is acknowledged to imply some degree of bias, best countered through a willingness to be surprised: "Despite our biases, we do in fact disconfirm our expectations all the time. Our ability to be surprised, to change our minds, to come to a new understanding, demonstrates that our initial biases are not immutable" [Sti93, p.613]. There is therefore an explicit acknowledgement that some degree of subjectivity is involved in all analysis.

Strauss and Corbin express a particular concern about bias, stressing the importance of *constant comparison* to ensure a genuine fit between data and the emerging theory: "We recognise the human element in analysis and the potential for possible distortion of meaning. That is why we feel that it is important that the analyst validate his or her interpretations through

constantly comparing one piece of data to another" [SC98, p.137].

Table 3.3 illustrated the epistemological differences between quantitative and qualitative approaches as reflected in methods chosen. It was also mentioned in Section 3.2.3.1 that grounded theory can be practised within either tradition [Cha95, Wil01]. Nonetheless, there is a further claim that criteria for evaluation will necessarily hinge on epistemological position. Willig proposes two versions of grounded theory, one that is *realist* and the other that is *constructionist* [Wil01]. The realist version is 'experiential' in that it aims to understand people's experiences; the constructionist version is 'discursive' in that it is more concerned with the role of language in the construction of reality. Thus the realist approach is concerned with the essence of *what* is said by respondents, rather than *how* it is expressed through subtle nuances in language.

This concern with the role of language links in with the researcher's theoretical view of whether the 'status of the text' is seen as a factual account of the respondent's experience, or as embedding hidden meaning. In the case of a constructionist version of grounded theory, the proposed criterion of evaluation is reflexivity. In the case of the realist version, Willig proposes objectivity and reliability as criteria for quality control. Objectivity is maintained by not imposing predefined categories on the data. Reliability refers to triangulation (of researchers and/or methods) to illustrate the convergence of analyses from different perspectives. The evaluation criteria proposed by Willig are summarised in Table 3.7.

Table 3.7: Different approaches to grounded theory and their impact on evaluation criteria (after Willig [Wil01])

Version of grounded theory	**Realist**	**Constructionist**
Slant	Experiential	Discursive
Concerned with	Essence of *what* is being said	*How* experience is being expressed through nuances in language
Evaluation criteria	• Objectivity (not imposing preconceptions on the data) • Reliability (triangulation of researchers or methods)	• Reflexivity (documenting how views changed during analysis)

All of the above criteria have been proposed in response to the challenging problem of evaluating qualitative analysis within a framework that openly acknowledges the subjective role of the researcher but ultimately prizes objectivity [Pid96]. No two investigators would

generate exactly the same names for concepts, and depending on their research questions they might choose to focus on different categories identified in the data. However, it was deemed important to ensure that the categories and interconnections in the analysis reported in this thesis seemed logical and convincing to others. The approach taken in terms of evaluation was to combine a selection of the criteria proposed above. First, evolving interpretations of the data were documented throughout the process *(reflexivity)*, demonstrating an attitude of permeability to the data. Secondly, each stage of the analysis was documented throughout to leave a sufficiently detailed *audit trail* for other researchers to validate the way the raw data was abstracted and interpreted. Finally, the results were verified with an independent researcher experienced with using qualitative research in psychology *(triangulation of researchers)*. The goal was not to agree on the actual concept and category names, but to validate the underlying logic of the interrelationships.

3.2.3.6 Departure from the method

There were two key ways in which the analysis reported in this thesis departed from the method as outlined by Strauss and Corbin. The first difference concerns theoretical sampling. Typically, grounded theory proposes a close interweaving between data collection and analysis, where ongoing analysis directly shapes further sampling decisions. In this particular case the coding procedures were applied post-hoc to a large body of relatively homogeneous data, all respondents having had a similar and controlled experience of interacting with an avatar. Willig refers to this as an 'abbreviated' approach to grounded theory, the 'full' approach being one where sampling decisions are iteratively shaped by analysis [Wil01]. Negative case analysis was therefore not applied as a criterium for evaluation, because no sampling occurred after analysis had begun [Pid96]. The second key difference was that a method of visual abstraction was developed to facilitate the final stages of analysis. It is common for researchers to develop their own systems of notation and diagrammatic representation while using grounded theory. This approach will be detailed with examples in Chapter 6.

3.3 Chapter Summary

The research described in this thesis consists of three experiments investigating different aspects of avatar fidelity. The first part of this chapter outlined aspects of experimental design and procedure common to all three. Although one experiment explores the use of psychophysiological data to investigate responses to virtual humans, the main emphasis is on subjective perceptions. Quantitative data from the questionnaires was analysed using logistic regression, a conservative method of statistical analysis that has the advantage of treating responses on a

Likert-type scales as ordinal, rather than interval, data. In addition, the interviews from the first experiment are analysed using a qualitative method of analysis called grounded theory. In introducing grounded theory, this chapter has touched on the epistemological debate that would pit quantitative and qualitative approaches against each other as representing opposing paradigms. It has also aimed to illustrate that the two approaches can be fruitfully combined, particularly in relatively new areas of research.

3.3.1 Reading guide to Chapters 4, 5 and 6

The experiments are reported in logical rather than chronological order. The experiment on responsiveness is reported in Chapter 4, though it was conducted between the two experiments on eye gaze described in Chapter 5. The findings from the grounded theory analysis of interview data from the first experiment will be reported in Chapter 6. Table 3.8 below is designed as a reading guide to the upcoming chapters. It details the main factors of interest in each experiment, along with the methods of data collection and analysis used in each case.

Table 3.8: Overview of the three experiments

Focus of experiment	**Responsiveness**	**Eye gaze**	**Eye gaze + Photorealism**
Logical order	Chapter 4	Section 5.1	Section 5.2
Chronological order	2	1	3
Location	UCL	BT Exact	UCL
Aspect(s) of avatar fidelity investigated	• Behavioural (responsiveness to proximity of participant)	• Behavioural (eye gaze)	• Behavioural (eye gaze) • Visual (appearance)
Responses analysed	• Questionnaire • Interview • Psychophysiological measures (Heart rate and EDA)	• Questionnaire • Interview (Analysed using grounded theory)	• Questionnaire • Interview

Chapter 4

Experiment on Responsiveness

The overarching goal of the three experiments described in this thesis was to explore the impact of avatar fidelity on social responses to virtual humans. As discussed in Section 2.4.4, there is evidence from previous research that people respond socially to agents with minimal behavioural cues, despite rationally knowing that they are not real people [BBBL01, PSB01a, PSB01b]. These findings are consistent with theory of Reeves and Nass [RN96] that people tend to anthropomorphise computer interfaces, responding to them as social entities. In the words of Biocca, "Humans may be hardwired to respond to affordances in the environment that are correlated with sentient beings such as other humans and animals. Mediated embodiments such as pictures, computer characters, moving robots, and other representations of 'apparently sentient' others may automatically trigger social presence responses. Social responses are triggered by representations that the user/viewer knows to be 'false', i.e. only a representation" [BH02, p.17].

The experiment presented in this chapter investigates the impact of minimal responsiveness, a subcomponent of fidelity that has received little explicit attention in the literature. In the context of collaboration in CVEs, perceptions of virtual humans will likely hinge not only on convincing animation, but also on their ability to respond appropriately to ongoing interaction. Heeter argues that responsiveness is an important factor in eliciting presence and social presence [Hee92]. Swinth and Blascovich propose that virtual humans can vary considerably in responsiveness [SB02]; they can range from being static and unresponsive to dynamic and responsive. The experiment presented in this chapter explored the lower boundaries of responsiveness by comparing responses to a group of virtual humans displaying incremental increases in animation and responsiveness. The goal was to investigate the degree to which participants would respond to them as social entities as their degree of dynamic and responsive behaviour increased across the conditions. Since the virtual humans in this experiment did not represent real people engaged in interaction, they will be referred to as agents rather than avatars.

Responses considered in this experiment include presence as well as different aspects of copresence, the degree of sentience attributed to the agents, and the degree to which participants modified their own behaviour to account for the agents' presence in the virtual room. In addition to subjective measures, heart rate and EDA were employed to gauge participants' objective responses to their experience in the IVE.

4.1 Experimental Aims and Expectations

The overall aim of the experiment was to investigate the degree to which agents were treated as social entities rather than objects as their responsiveness increased across the four conditions. The basic hypothesis was that the greater the responsiveness of the agents, the higher the likelihood that participants would have had an experience of being with people, measured on a number of indicators.

4.2 Experimental Design

A between-groups one-way design was employed with the main four-level factor being the degree of agent responsiveness. The scenario involved exploring a virtual library that contained five virtual 'readers' seated around a central table, as illustrated in Figure 4.1.

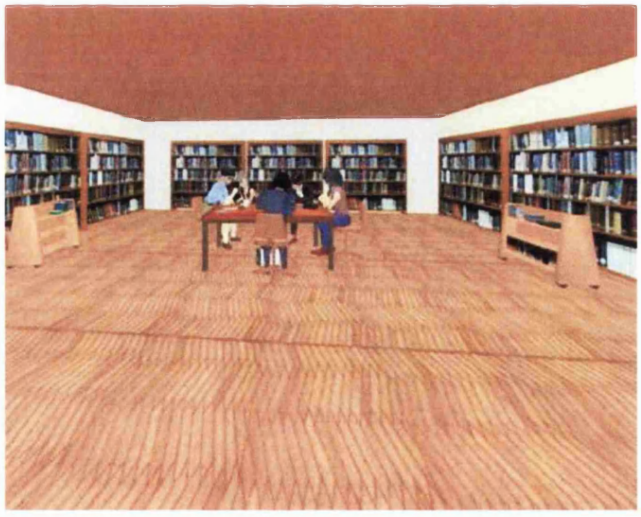


Figure 4.1: Virtual 'readers' as seen upon entering the virtual library environment

4.2.1 Independent variables

The four conditions are listed in order of increasing responsiveness.

Condition 1 (Static): Agents were static, frozen in a reading pose.

Condition 2 (Moving): Agents were animated, carrying out such behaviours as might nor-

mally be expected of people in a reading room, including fidgeting, turning pages and occasionally looking around. They did not respond to the participant's position in the room.

Condition 3 (Responsive): The behaviours of condition 2 (*moving*) were supplemented by agent responsiveness to the participant's location in the space. When the participant approached, each agent would respond by changing posture and engaging in gaze behaviour (see Figure 4.3). The rules for responding to the participant's position were based on four 'interpersonal zones' separated by a series of nested circles. Hall's [Hal66] model for interpersonal distance outlines the appropriate distance for different types of social interaction: intimate, personal, social-consultative and public (Figure 4.2). A similar logic was applied to the agents' reaction zones, although the distances were modified during the pilots to make them more effective for the Cave (Table 4.1).

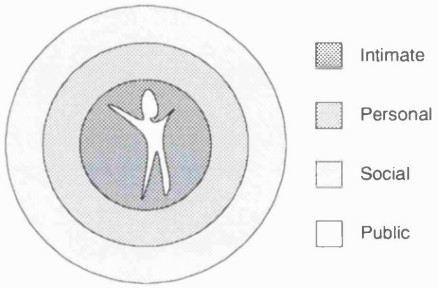


Figure 4.2: Response zones for agents in conditions 3 *(Responsive)* and 4 *(Talking)*

Table 4.1: Comparison of Hall's [Hal66] suggested measures of interpersonal distance and those adapted for use in the Cave

Interpersonal Zone	**Hall's suggested measurements (metres)**	**Measurements adapted for Cave (metres)**
Intimate	0 - 0.5	0 - 1.5
Personal	0.5 - 1.22	1.5 - 2.44
Social	1.22 - 3.65	2.44 - 3
Public	over 3.65	over 3

When the participant was in the agent's 'public' zone, the agent would move as in condition 2 (*moving*), focusing on study-related things around the table. As the participant moved through each zone towards the agent, the agent's behaviour would change as follows: gaze would increase, posture would become more upright, the frequency of gaze and posture shifts would increase, and the probability that the agent would look at the user would increase. Finally, in the *intimate* space, the agent would physically turn around in its chair and visually

track the participant until interpersonal distance was increased. The degree of each agent's visual engagement with the participant was therefore a function of interpersonal distance.

Condition 4 (Talking): In this condition the responsive behaviours were the same as in condition 3 except that the first agent approached would speak to the participant briefly. The verbal content was in a language not recognisable to participants, but the tonality suggested a question, followed by a pause, then another question. After a few seconds, the agent would say 'O.K.' as if resigned to not getting an understandable answer, and would turn back to its activities. The purpose of these utterances was to suggest that verbal communication was possible in principle, though not in practice due to the language barrier.

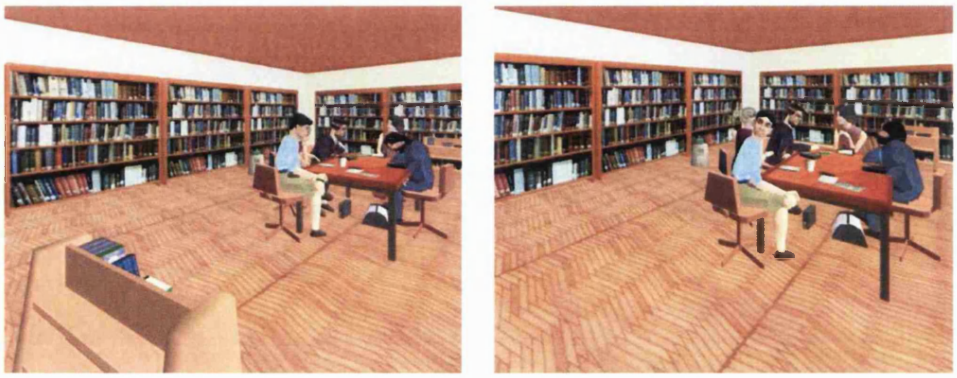


Figure 4.3: Virtual readers in the library room would visually respond to the participant's proximity in the *responsive* and *talking* conditions.

4.2.2 Piloting

This was an exploratory experiment and a significant emphasis was placed on piloting. Twenty-six pilot sessions were conducted over a six-month timespan with members of the Department of Computer Science at UCL. During the piloting phase numerous adjustments were made to the experimental design based on feedback from participants. Since the aim was to investigate responses to virtual humans in the absence of sustained verbal communication, a neutral scenario was needed that would explain the presence of the virtual humans without raising expectations for direct interaction or verbal communication. Both the scenario and task changed considerably as a result of comments from pilot participants.

4.2.2.1 Piloting the task

A task was needed that would enable participants to spend time in the virtual library and notice the agents, without needing to explicitly interact with them. A search task seemed to offer a convincing pretext to keep participants in the virtual library for several minutes. At first a very specific scenario was designed involving a health and safety inspection. The problem with

this task was that looking for flaws in the space could lead participants to an overly critical scrutiny of everything in it, including the agents. This task was abandoned in favour of a more neutral search task for a mobile phone someone had left in the room. This in turn presented two problems: firstly, it was extremely challenging to make the VE appear sufficiently cluttered to hide a small object. Despite efforts to fill the room with objects on various surfaces, the mobile remained easy to find, allowing participants to complete the task too quickly. In order to maximise consistency between experimental sessions it was preferable for all participants to spend the same amount of time in the VE.

Several sessions were piloted in which participants were asked to search for a smaller object such as a watch or pair of keys that someone "may have" left in the library. To prevent participants from completing the task too quickly, the object in question was not actually placed in the VE, and when the requisite five minutes had passed they were thanked for their efforts and informed that it had been found elsewhere. This did not entirely solve the problem of keeping session times constant, because a few individuals gave up when they felt they had looked thoroughly for the object without finding it. It also did not seem satisfactory, for ethical reasons, to deliberately mislead participants into a potentially frustrating search for an object that was not in the space. The search task also led participants to explore the space selectively by looking only in certain likely areas. In the words of one pilot participant, "If I'm looking for a watch, I'll look in 'watchy' type places: not in books, but under tables." Also, some participants were so intent on searching for the object that they paid little attention to the agents, making it impossible to ensure that they had experienced the condition as intended.

A simple exploration task was piloted. At first, participants were asked to look around the room for a certain amount of time and then draw a detailed plan of what they had seen. However, memorising the space again meant exploring it with a selective goal in mind, and drawing the plan took up a disproportionate amount of time in the procedure without yielding useful data. A final decision was made to keep the task as simple as possible, with participants being given a finite amount of time to thoroughly explore the space so that they could later report on what they had experienced.

Task time needed to be sufficient to allow participants to observe the environment as well as the agents, but not so long that they risked developing simulator sickness. Pilot sessions were conducted with a five-minute and ten-minute exploration time, and it was finally determined that four minutes was optimal for thoroughly exploring the space while minimising the total amount of time spent in the Cave.

4.2.2.2 Piloting the scenario

A library setting seemed to provide a familiar context where social conventions dictate that interaction is possible but not commonplace. In a library it would appear natural for the agents to be in the room without any explicit verbal interaction taking place. A virtual library model was constructed containing study tables, bookshelves, book trolleys and other recognisable objects. In the initial version, the square library environment was divided into four equal quadrants, each containing a study table. The agents were seated at one of four tables, with a view to investigating whether participants would tend to approach or avoid the 'social area' in the VE.

However, this idea was abandoned for three reasons. Firstly, the placement of the agents in one of the four quadrants would need to be randomised to exclude variables such as participant handedness influencing spatial behaviour. Secondly, the table with the agents seated around it necessarily occupied more floor space in its own quadrant, meaning that participants might tend to avoid it because of crowding, and not on account of a social response to the agents. Finally, several participants mentioned that the library, with its four tables, was too cluttered and left them insufficient room to navigate; their desire to avoid colliding with objects and agents meant that they were forced to move in a narrow space and make frequent rotations, increasing their sense of simulator sickness. A final decision was therefore made to seat the agents around a central table.

4.2.2.3 Piloting the pre-questionnaires

Being computer scientists, many of the pilot participants were experienced programmers. Several reported thinking of the agents as part of a computer 'system' rather than as people. A question regarding technical literacy was therefore included in the pre-questionnaire as an additional explanatory variable. Another key explanatory measure captured in the pre-questionnaires concerned participants' social anxiety in everyday life. The standard Social Avoidance and Distress (SAD) questionnaire [WF69] contained explicit questions about social attitudes, increasing the likelihood of demand characteristics that would undo any effort to design a neutral scenario and task. The questionnaire therefore included a number of questions about reading, books and work environments in an attempt to deflect attention from the social content of the questionnaire. Appendix B contains the modified questionnaire.

4.2.2.4 Piloting the post-test questionnaire

A number of questionnaire items were created for each of the variables of interest in the experiment. In the debriefing sessions participants explained how they had interpreted the questionnaire items. The order and phrasing of the items changed considerably in response to this

feedback. For example, the agents were referred to as 'virtual characters' or 'readers' to avoid unfamiliar jargon.

4.2.2.5 Piloting the agents

One important outcome of the pilots concerned the pose of the agents in the *static* condition. Originally, all five agents were posed in an upright, seated position. However, responses in the pilot sessions indicated that they appeared too robotic. The agents were subsequently posed in a position that suggested reading or studying, in order to create the idea of activity even in the absence of animation.

4.2.2.6 Piloting the procedure

Several pilot participants reported a sense of discomfort in colliding accidentally with objects and agents. This was not deliberate, rather a result of lack of experience with navigating around the IVE. It was therefore decided to include a brief training task in the procedure. An adjacent training space was constructed next to the library so that participants could practise moving in the IVE before entering the virtual library. The need for a training task made it especially important to limit the main exploration task to four minutes, in order to ensure that total time spent in the Cave did not exceed 10-15 minutes. Again, this was in order to minimise the risk of simulator sickness.

Participants mentioned that any noise they heard distracted from the experience. For this reason, special care was taken to keep absolutely quiet in the laboratory during the exploration task.

4.2.3 Apparatus

The laboratory consisted of two separate spaces: a reception room where participants completed questionnaires, and a laboratory containing the Cave and the PC for monitoring participants' psychophysiological responses.

4.2.3.1 ReaCTor

The Cave-like system used was a ReaCTor made by Trimension, consisting of three 3m x 2.2m walls and a 3m x 3m floor. It is powered by a Silicon Graphics Onyx2 with 8 300MHz R12000 MIPS processors, 8GB RAM and 4 Infinite Reality2 graphics pipes. The participants wore CrystalEyes stereo glasses which are tracked by an Intersense IS900 system. They held a navigation device with 4 buttons and an analogue 3D mouse that is similarly tracked; all buttons except for the 3D mouse were disabled to stop participants from manipulating objects in the VE. The 3D mouse was used to move around the VE, with pointing direction determining the direction of movement enabled for the horizontal plane only.

The software for this VE was written on top of the Distributed Interactive Virtual Environment (DIVE) software platform [FSS+01]. DIVE is an internet-based multi-user virtual reality system in which participants can navigate in a shared 3D space and interact with each other. The software used was implemented on a derivative of DIVE 3.3x [FSS+01]. This was recently ported to support spatially immersive systems [MVS+02]. Since DIVE also supports the import and export of VRML and several other 3D file formats, it was possible to import ready-made avatars from other projects [PSB01a, PSB01b, SHS+00].

4.2.3.2 Psychophysiological monitoring equipment

Participants were fitted with Thought Technologies Ltd. ProComp+ EKG sensors on their torso and EDA sensors on their non-dominant hand [Tho03]. The software for visualising and recording these measures was run on a dual-processor Dell PC. The ProComp box was kept securely in a pack strapped to the participant's waist and linked to the PC via serial cable (see Appendix G for images of the monitoring equipment).

4.2.4 Agents

Three male and two female agents were used as shown in Figures 4.3. The virtual human models were originally from Criterion Software's RenderWare product. Their appearance was edited and the body parts segmented and arranged into a hierarchy. Their original animations included walking, standing, sitting and pointing gestures. Additional behaviours such as blinking and leaning forward over books were implemented in DIVE as Tcl scripts that controlled the relevant parts of the body hierarchy.

4.2.4.1 Agent animations

In each of the four conditions, each agent's state had two main features: gaze and postural behaviour. Gaze behaviour for each agent was broken down into three components. The first was the list of objects it could look at (including the books in front of it, other agents and the human participant in the experiment). The second was the probability that it would look at any one object; for instance, the probability of looking at the participant ranges from 0 when in the public space to 1 when in the intimate space (Table 4.2). The third was the frequency with which it changed its visual focus; this frequency reduced dramatically in the agent's intimate space, where the agent effectively stared at the participant.

Postural behaviour consisted of two factors: the repertoire of poses the agent could adopt and the frequency with which it changed between poses. When the agent changed its pose, a new pose was randomly selected from the list of poses available for the current interpersonal zone. The list of available poses for the *intimate* and *personal* zones contained more 'upright'

Table 4.2: Probability and frequency of agent gaze at participant for each interpersonal zone

Interpersonal Zone	Probability of agent looking at participant
Intimate	1
Personal	0.7
Social	0.5
Public	0

poses, denoting alertness to the participant.

4.2.4.2 Agent audio for the talking condition

Four sound files were recorded in Turkish language, two in a male and two in a female voice. Each sound file consisted of a brief phrase with an upward inflection, suggesting a question. The actual meaning of the phrases was, "Can I help you," and "Are you looking for something." Again, the purpose of recording the utterances in a language other than English was to make conversation seem possible in principle, but not in practice. Special care was taken that participants taking part in this condition neither spoke nor understood Turkish. Depending on the gender of the first agent to be approached, either the male or female sounds were played, with a slight pause between them suggesting that the agent was waiting for a response from the participant.

4.2.5 Virtual environment

The VE consisted of a virtual 'training' room and and adjacent 'library' room (see Figure 4.4). The training space was designed to help participants become accustomed to navigation through the VE. Large single-digit numbers from 1 to 9 were scattered around the training room, and participants were instructed to move through these in numerical order. The doorway separating the virtual rooms remained closed throughout the training period.

The objects in the model were made non-graspable to prevent participants from lifting objects or agents with the 3D mouse. Collision detection was enabled on the walls to prevent participants from straying from the library during the experiment, or to glimpse the library space during the training period. However, there was no collision detection on the objects or agents themselves, so participants were free to walk through them. This was done deliberately to see whether participants would show the same tendency observed in many pilots of avoiding collision with objects. One aim was to explore whether, as in Bailenson et al.'s proxemics experiment [BBBL01], people would show a different concern for characters than for virtual objects.

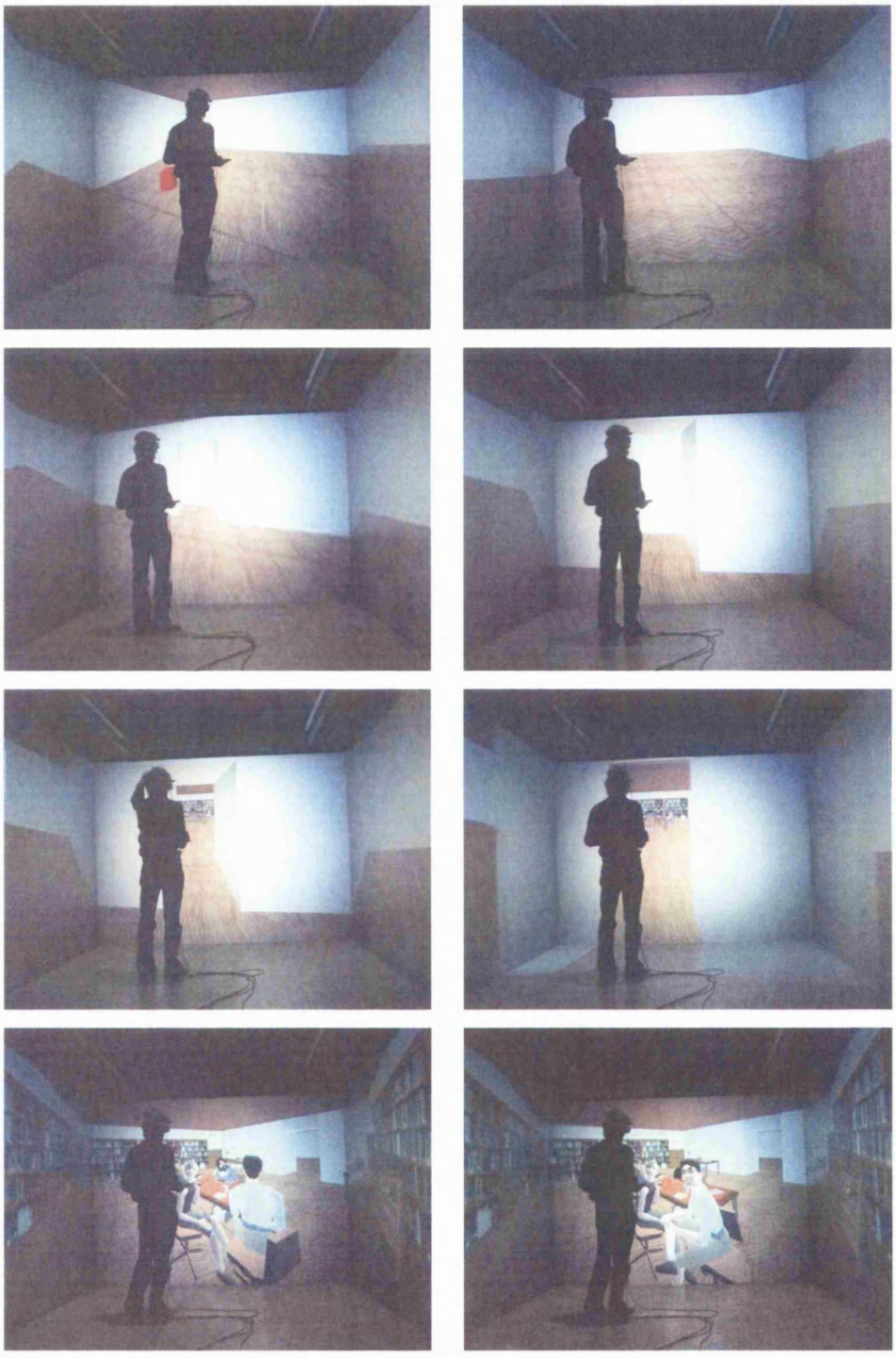


Figure 4.4: Images showing participant in the virtual training room, then moving through the doorway to approach the characters in the virtual library

4.2.6 Task

Participants were instructed as follows for the exploration task:

> "In a few moments the doorway will open. Please go through the doorway and you will find yourself in another room. When you are in there, please observe your surroundings. Afterwards we will be asking you about what you experienced."

They were told they had four minutes in the virtual library room, and that when the time was up, they should return to the (virtual) training room, at which point the experimenters would return to remove all equipment and sensors. No mention was made of the characters, and no further instructions were given as to how to explore or observe the space.

4.2.7 Population

Participants were recruited from the campus through an advertising poster campaign. They were paid £5 for the one-hour study. As far as possible a gender balance was maintained across the conditions, as shown in Table 4.3:

Table 4.3: Factorial design for the experiment on responsiveness

Condition	Male	Female	Total
Static	7	4	11
Moving	6	4	10
Responsive	6	4	10
Talking	5	5	10
Total	24	17	41

There were 11 participants in condition 1 because one additional participant was recruited due to an earlier mistaken belief about missing data. All data collected was then used in the analysis.

4.2.8 Procedure

The procedure for this experiment was divided into three phases: one prior to the experience in the Cave, one including the training and exploration task in the Cave, and one after it. Two experimenters collaborated on the running of each session. The author administered the questionnaires, led participants through each phase of the experiment, and conducted the final interview. A collaborator oversaw the psychophysiological monitoring. Sessions for each of the four conditions were interspersed to avoid the experimenters' habituation to the procedure affecting the way the experiment was run across the conditions. Each session was run in three phases.

4.2.8.1 Phase 1: before entering the Cave

Participants were welcomed in a reception area and were given an information sheet and consent form (see Appendix B). They were left in private to attach the EKG sensors to the chest area. Next, the EDA sensors were attached to their non-dominant hand and all the equipment was placed inside a pack fastened to their waist (Figure G.3). The second experimenter ensured that the monitoring equipment was transmitting a clear signal, and the 'at rest' baseline was recorded while participants filled out the standardised Social Avoidance and Distress (SAD) questionnaire [WF69].

4.2.8.2 Phase 2: inside the Cave

Upon entering the Cave, participants were asked to stand in the centre and observe the double image projected on the walls; this was to ensure that they could confirm they were seeing in stereo when wearing the stereoscopic goggles. They were then equipped with goggles and a head tracker and were asked to take a few paces around the space and bend down, to experience the way their viewpoint within the VE responded to their height and position in space.

Since the virtual space was larger than the physical space of the Cave, they were then shown how to navigate around the space using the 3D mouse. Participants then carried out a brief training exercise designed to increase their confidence with navigation, to minimise unwanted collisions with virtual objects. Once participants felt comfortable, they were reminded that the session would be stopped immediately if they felt any nausea. They were asked to stand in the centre of the Cave facing the front wall, and to position themselves directly in front of the virtual doorway, that was still closed. At this point the task was read out to them and they were reminded that the experimenters would be able to hear them but would not be speaking to them. A curtain was then drawn to visually isolate the Cave from the rest of the lab.

The task began when the virtual doorway was opened by the experimenter. During the task, the second experimenter marked the psychophysiological data stream for events of interest including collisions with agents and objects, participants moving close to agents, and participants trying to interact verbally or noverbally with agents (for instance by waving). After four minutes in the virtual library room, participants were asked to return to the virtual training space in order to transition to the real world. Immediately afterwards they were led through to the reception area.

4.2.8.3 Phase 3: after leaving the Cave

Participants filled out a post-questionnaire about their sense of presence and their responses to the virtual characters. The session concluded with a semi-structured interview (the question-

naire and semi-structured interview agenda are contained in Appendix B).

4.2.9 Response variables

A number of different response variables were considered in the analysis, including *copresence, presence, participant behaviour* and *perceived awareness* of the agents. The variables were constructed from n questionnaire items, each on a 1 to 7 scale with the score adjusted for analysis so that the higher score represented a higher response. The items for each response are detailed below.

1. **Copresence** (n= 5): This is the extent to which the participants had the sense of being with other people.
   (a) During the course of the experience, did you have a sense that you were in the room with other people or did you have a sense of being alone?
   (b) Now consider your response *over the course of the whole experience.* To what extent did you respond to them [the characters] as if they were people?
   (c) To what extent did you have a sense of being in the same space as the characters?
   (d) Now consider your response *over the course of the whole experience.* To what extent did you have a sense of personal contact with the characters?
   (e) Now consider your response *over the course of the whole experience.* Did you respond to the characters more the way you would respond to people, or the way you would respond to a computer interface?
2. **Presence** (n= 5): Information was collected on the sense of presence in the virtual environment using the set of five questions from the SUS questionnaire [SSU93]. In these questions, 'room' refers to the VE, whereas 'laboratory' refers to the physical laboratory.
   (a) Please rate *your sense of being in the room,* where 7 represents your *normal experience of being in a place.*
   (b) To what extent were there times during the experience when the room was the reality for you and you almost forgot about the real world of the laboratory where the experience was really taking place?
   (c) When you think back about your experience, do you think of the room more as *images that you saw,* or more as *somewhere that you visited*?
   (d) During the course of the experience, which was strongest on the whole, your sense of being in the room, or of being in the real world of the laboratory?

(e) During the time of the experience, did you often think to yourself that you were just standing in a laboratory or did the room overwhelm you?

3. **Participant behaviour** (n= 5): The extent to which participants reported that they altered their behaviour in response to the agents.

   (a) How far did you make an effort to avoid disturbing the characters?

   (b) How far did you feel inhibited in your task by the characters?

   (c) Now consider your response *over the course of the whole experience.* How much did you want to interact with them?

   (d) Did you attempt to interact with them?

   (e) To what extent did the presence of the characters affect the way you explored the space?

4. **Perceived agent awareness** (n= 4): The extent to which the participants perceived that the agents were aware of them.

   (a) How much did the characters seem to respond to you?

   (b) How much were the characters looking at you?

   (c) How much did the characters seem aware of you?

   (d) To what extent did you feel observed by the characters?

Objective response variables: An attempt was made to record each participant's heart rate EDA. Readings were taken throughout participants' time in the virtual training and library room.

4.2.9.1 Explanatory variables

Explanatory variables included gender and computer usage ('the extent to which you use a computer in your daily activities') on a 1 to 7 scale (1 = Not at all, and 7 = Almost all the time). Also, participants' degree of social anxiety in everyday life was measured using Social Avoidance and Distress (SAD) questionnaire [WF69]. The social anxiety score was taken into consideration in relation to participants' self-reported social responses, such as their reported tendency to approach or avoid the agents.

4.3 Results

This section begins by presenting the findings for the subjective response variables detailed above, and concludes with the analysis of the psychophysiological responses (heart rate and EDA).

Table 4.4: Mean ± standard deviations of count response variables

N = number of participants in each condition.

n = number of questions on which the count is based.

	Static (N=11)	Moving (N=10)	Responsive (N=10)	Talking (N=10)
Copresence (n=3)	0.55 ± 0.69	1.10 ± 0.88	1.50 ± 1.90	1.20 ± 1.14
Presence (n=5)	1.64 ± 1.43	1.90 ± 1.60	1.00 ± 1.70	2.00 ± 1.83
Participant behaviour (n=5)	0.73 ± 0.90	1.20 ± 1.14	1.30 ± 1.16	1.90 ± 0.74
Perceived agent awareness (n=4)	0.09 ± 0.30	0.10 ± 0.32	2.10 ± 1.60	2.10 ± 1.37

Table 4.4 shows the mean and standard deviations of the counts of 'high' responses (6 or 7) across the n questions in each condition. Figure 4.5 shows the means of the raw questionnaire responses, and illustrates a progressive increase in mean responses from the *static* through to the *talking* condition for both *copresence* and perceived agent *awareness*. The method of logistic regression described in Section 3.2.1 was used to check for significance, and also to test whether other exogenous variables should be included in the model. The findings for each response variable are presented below.

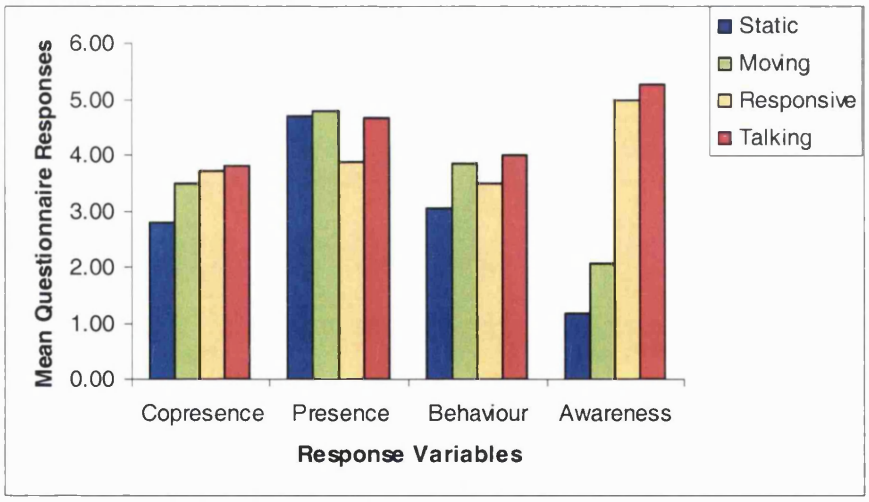


Figure 4.5: Means of raw questionnaire responses

4.3.1 Copresence

There is no significant difference between the conditions for the main effect on *copresence* when considered by itself. However, introduction of the variable 'computer usage' (the extent of using a computer in everyday activities) as a covariate changes this picture. When this is brought into the analysis, then condition 3 (*responsive*) results in the highest response (this is significant at the 5% level). The greater the degree of computer usage, the less participants

reported a sense of copresence with the agents. The change in deviance (Δdeviance) due to the interaction term is 13.68. This has an approximate χ^2 distribution on 3 d.f. (see Section 3.2.1). This is significant at p=0.0033.

4.3.2 Presence

Condition is not significant when considered alone. However, as with copresence, introducing the variable 'computer usage' as a covariate changes the picture. Condition 4 *(talking)* results in a significantly higher response. Again, the greater the degree of computer usage, the less likely participants were to feel present (Δdeviance=19.32, d.f.=3, p=0.0002). The results also indicate a significant positive correlation between reported presence and copresence (Δdeviance=10.53, d.f.=1, p=0.0145).

4.3.3 Participant behaviour

This variable focuses on the impact of the agents' presence had on participants' self-reported behaviour. There is a significant difference between the conditions for the main effect on *participant behaviour* (Δdeviance=9.42, d.f.=3, p=0.024). In this case condition 4 (*talking*) is significantly higher than the *static* condition, and none of the other three conditions are significantly different from each other. Interestingly in this case, there is no 'computer usage' effect. Looking at the subcomponents, question 3(c) concerning the desire to interact with the agents follows the same pattern, where only condition 4 (*talking*) is significant. For question 3(d) concerning the attempt to interact with the agents, all three non-static conditions (2, 3 and 4) are significantly higher than the *static* condition (1), but not significantly different from each other.

There is a significant positive correlation between question 3(a) (concerning the attempt to avoid disturbing the virtual characters) and the SAD score (r = 0.55, t = 4.1 on 39 d.f.). This is evidence that participants acted towards the virtual characters in a way that might be predicted from their social anxiety score (i.e., following what might be expected in their everyday behaviour). However, such a correlation is not shown with the other subcomponents of participant behaviour.

4.3.4 Perceived agent awareness

This variable concerns the extent to which the participants perceive the agents to be aware of them in various ways. Table 4.5 shows the results of the regression analysis. Conditions 3 (*responsive*) and 4 (*talking*) are significantly higher than conditions 1 (*static*) and 2 (*moving*). This is as expected, because objectively in these conditions the agents are are visually responsive to participants, denoting 'awareness' of their movements in the virtual library. Once again there is no effect of 'computer usage'. The same significant result is found for each of the four

subcomponents of this response.

Table 4.5: Logistic regression with *perceived agent awareness* as the response variable

Overall deviance = 70.53, d.f. = 5

The deviance column shows the increase in deviance that would result if the corresponding variable were deleted from the model.

Variable	Deviance χ^2	d.f.	5% χ^2
Condition	61.11	3	7.815
Gender (female)	8.22	1	3.841
VR experience	4.02 (+)	1	3.841

In addition, both *gender* and previous *VR experience* are significant. Women were more likely to perceive the agents as being aware; the same is true for participants with prior experience with VR systems.

4.3.5 Analysis of objective responses

Physiological data was not available for all participants, because the equipment did not always function correctly. Data was collected during the time each participant was in the virtual training room and virtual library room. As in Meehan's analysis [MIWB01], the analysis reported here only considers overall means. The response variable of interest for both heart rate and EDA is the change from the overall mean level in the training room to the overall mean level in the library room.

Table 4.6: Mean $\pm$ standard deviations of heart rate response (beats per minute)

N = number of participants in each condition.

Condition	*N*	Training room: Mean $\pm$ standard deviations	Library room: Mean $\pm$ standard deviations	p
Static	8	75.40 $\pm$ 11.41	77.36 $\pm$ 12.40	0.450
Moving	7	80.37 $\pm$ 11.64	82.42 $\pm$ 12.33	0.216
Responsive	8	58.65 $\pm$ 17.88	64.93 $\pm$ 17.36	0.034
Talking	10	76.09 $\pm$ 10.39	79.49 $\pm$ 8.82	0.124

The results shown in Tables 4.6 and 4.7 illustrate that both heart rate and EDA increased when participants went into the richer library environment from the training room. For EDA this difference is significant for all conditions, with the exception of condition 2 (*moving*), where it is just below significant. For heart rate the difference between the mean levels in the training room and library room is only significant for participants in condition 3 (*responsive*).

Using *presence* and *computer usage* as covariates indicates that this heart rate increase

Table 4.7: Mean ± standard deviations of EDA response (milliSiemens)

N = number of participants in each condition.

Condition	*N*	Training room: Mean ± standard deviations	Library room: Mean ± standard deviations	p
Static	8	7.64 ± 2.19	8.89 ± 2.65	0.008
Moving	7	10.89 ± 4.20	12.42 ± 5.43	0.064
Responsive	8	11.37 ± 5.52	13.64 ± 6.05	0.003
Talking	10	10.99 ± 5.50	13.13 ± 7.63	0.034

in condition 3 diminishes with computer use, and also diminishes with increasing reported presence in the library room. The decrease in heart rate with increasing computer usage is consistent with the findings for *copresence* and *presence* reported above.

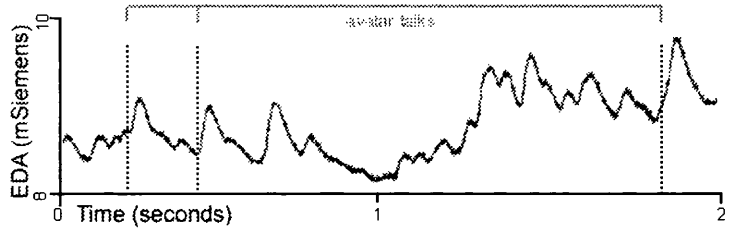


Figure 4.6: EDA reading for random participant in condition 4 (*talking*). Dotted lines, indicating when an agent 'talks,' are followed by a sharp increase in EDA.

An experimenter marked significant events in the data stream during the time participants were in the library room A participant is selected at random from condition 4 and the graph of their EDA reading is shown in Figure 4.6. This indicates the concomitant sudden jump in EDA for such significant events as the avatar speaking. Although data on the individual events was not used in the analysis, it pointed to directions for future research that will be discussed in Chapter 7.

4.4 Chapter Summary

The experiment presented in this chapter aimed to isolate agent behaviours from verbal interaction to understand how responses can be affected by simple increments in agent responsiveness. The goal was to study participants' responses in an unstructured social context where they were free to explore a virtual room without being explicitly instructed to interact with the agents.

None of the participants reported responding to the agents as if they were 'really' people. However, many expressed surprise at the fact that they had respected some social norms despite the fact that they knew the agents were computer-generated. People with higher levels of social

anxiety were significantly more likely to avoid disturbing the agents. This supports previous findings that on some level people can respond to agents as social actors even in the absence of two-way verbal interaction.

The analysis compared people's responses to visually identical agents whose behaviours were modified to reflect increasing levels of responsiveness. Results indicate that participants who encountered the visually responsive agents in condition 3 experienced a significantly higher sense of copresence. This effect diminished with experienced computer users. Interestingly, this pattern was precisely reflected in the heart rate data, suggesting a parallel between the subjective and objective responses.

The findings support the hypothesis that increasing agent responsiveness even on a simple level can impact on certain aspects of people's social responses. However, there were a variety of factors at play, including the behaviours of the agents, people's prior experience with technology and their responses to real-life social situations. These points will be discussed in Chapter 7.

Chapter 5

Experiments on Eye Gaze and Photorealism

This chapter presents two closely related experiments investigating two aspects of avatar fidelity: behavioural fidelity in terms of eye gaze, and visual fidelity in terms of photorealism. Where Chapter 4 explored social responses to agents in the absence of sustained verbal communication, these experiments investigate the avatar's perceived contribution to conversation between two human participants. Some of the challenges involved in developing expressive avatars were discussed in Section 2.4. Research on nonverbal behaviour in face-to-face communication [Arg88] can offer valuable leads on how to improve avatar expressiveness without resorting to full tracking.

One goal shared by both the experiments discussed in this chapter was to investigate the potential contribution of minimal behavioural cues to the perceived quality of dyadic communication. A single behaviour is selected for investigation: eye gaze. As discussed in Section 2.3.3.2, gaze it is a richly informative behaviour in face-to-face interaction. It serves at least five distinct communicative functions [AC76, Ken67]: regulating conversation flow, providing feedback, communicating emotional information and the nature of interpersonal relationships, and avoiding distraction by restricting visual input.

The experiment presented in Section 5.1 was designed to investigate the impact of eye gaze on perceived quality of communication. An avatar with random head and eye movements was compared to a visually identical avatar that combined simple head tracking with 'while speaking' and 'while listening' eye animations inferred from the audio stream. The design of these eye animations was informed by social psychology research on the differences in gaze patterns while speaking and while listening in face-to-face interaction [AC76, AI72, Ken67]. Both avatar conditions were compared to video and audio-only baseline conditions.

The experiment presented in Section 5.2 was designed to directly extend the research presented in Section 5.1. Where the first experiment was conducted in a non-immersive setting, the second assessed the impact of the inferred-gaze and random-gaze animations in a more de-

manding immersive setting where participants were free navigate in the 3D IVE. A second core question addressed in this experiment concerns the relationship between visual and behavioural fidelity. As discussed in Section 2.4.3, one assumption made by several researchers is that convincing behaviour is a higher priority than realistic appearance in the development of expressive avatars. This experiment uses a two-by-two factorial design to investigate both behavioural fidelity (in terms of eye gaze), and visual fidelity (in terms of photorealism). It addresses the question of whether the impact of animation is independent of the avatar's appearance.

This chapter will begin by addressing some related studies on eye gaze in mediated communication. The description of the two eye gaze experiments will follow.

5.0.1 Research on the role and function of gaze

Research on gaze in mediated communication has been concerned mainly with issues of conversation management in multiparty interaction. One of the perceived limitations of telephony-based videoconferencing systems is that they do not support selective gaze [Bux92, Sel95, Ver99]. Various media space systems have attempted to address this limitation by distributing individual audiovisual units in physical space to represent each user (see [Ver99] for a review).

Studies in CVEs have attempted to address the problem of how to support selective gaze in multiparty interaction within a shared 3D space. The GAZE groupware system [Ver99] is designed to ease turn taking by conveying gaze direction in a shared virtual space using VRML2. This system uses an advanced desk-mounted eyetracking system to measure where each person is looking. The gaze information is then represented metaphorically in the form of a 2D texture-mapped 'persona' which moves about its own x- and y-axis in the 3D environment.

Taylor and Rowe [TR00] argue that the GAZE groupware system is problematic for two reasons. First, using a snapshot instead of video precludes any possibility of expressing other nonverbal cues through the persona. Second, the use of a plane makes it difficult to generate the kinds of profile views useful in multiparty communication. They address these limitations by rendering video of the facial region on a generic 3D model of a face. Their system animates the head movement by tracking the two earphones and microphone to obtain head position information for each user. The eye movement is contained in the video image. Their system renders avatars from an asymmetric viewpoint that corresponds to the position of the real participant, who typically sits 20 inches away from a 14-inch desktop screen. They conclude that this system improves group interaction by preserving the semantic significance of gaze. However, integrating video as a part of gaze animation fails to address the needs of users who prefer to remain visually anonymous behind a synthetic avatar.

Both of the above studies are concerned with supporting selective gaze in groups of three

or more. In terms of two-person communication, Colburn, Cohen and Drucker [CCD00] present findings from an experiment comparing visual attention to the screen during 20 dyadic conversations using an avatar. Participants were presented with three 3-minute visual stimuli in random order: a blank screen, a fixed-gaze avatar and an avatar with a functioning eye gaze model, based on who was speaking and whether or not the participant was looking at the screen. Participants looked at the screen more when the avatar was present and most of all when the gaze model was active. The experiments presented in this chapter extend this research by investigating the impact of eye animations on a range of subjective responses including perceived quality of communication, presence and copresence.

5.1 Experiment on Eye Gaze

This section presents an experiment designed to investigate the importance of eye gaze in humanoid avatars representing people engaged in conversation. The experiment was conducted using a video-tunnel setup and was therefore not immersive. It compared responses to dyadic conversation in four mediated conditions. An avatar with 'random' head and eye movements was compared to a visually identical 'inferred-gaze' avatar that combined simple head tracking with 'while speaking' and 'while listening' eye animations inferred from the audio stream. The design of these eye animations was informed by social psychology research on the differences in gaze patterns while speaking and while listening in face-to-face interaction [AI72, AC76, Ken67]. Both avatar conditions were then compared to video (with audio) and audio-only baseline conditions. The impact of each condition on the quality of communication was assessed by comparing participants' subjective responses along four dimensions: how natural the conversation felt, their degree of involvement in the conversation, their sense of copresence, and positive or negative evaluation of the conversation partner. Additional subjective responses were gathered through interviews, and findings from the grounded theory analysis of transcripts from the two avatar conditions are reported in Chapter 6.

5.1.1 Experimental aims and expectations

The purpose of including the video and audio-only baseline conditions was to investigate where an avatar-mediated interaction would fall on a continuum of mediated communication ranging from having no visual image to having a high-quality, real-time video image of the conversation partner. The goal of the experiment was therefore two-fold: firstly, to test whether an avatar could contribute to the perceived quality of communication given minimal behaviours. The second, more specific goal was to examine the role of gaze: when the avatar's gaze was directly related to the conversation, would this improve the quality of communication compared to the

visually identical random-gaze avatar.

The expectation was that the *inferred-gaze* condition would lead to an improvement in perceived communication quality over audio-only. It was also expected that the inferred-gaze avatar would outperform the random-gaze avatar, based on the logic that its eye movements were related to an aspect of the conversation taking place. However, it was not known whether having an avatar with random gaze would be better or worse than not having an avatar at all. The video condition was expected to always outperform the other conditions in terms of perceived quality of communication.

5.1.2 Experimental design

100 participants were randomly assigned to one of four conditions representing different methods of mediated communication. The conversations took place over a video-tunnel link. Video tunnels are designed to correct for camera offset by offering a face-on view of the conversation partner [Bux92]. A deliberate choice was made not to make use of the 3D potential of the avatar, in order to abstract away from everything but the presence of the avatar and the effect of the head and eye animations. Therefore, only a head-and-shoulders view of each avatar was displayed. Each group was randomly divided into pairs of participants. The two people in each pair did not know each other. They were of the same gender, and were matched approximately for age (20s, 30s, 40s or 50s). They performed a ten-minute role-playing task in which they were randomly assigned to play one of two roles–a mayor or a baker. A detailed description of the design and running of the experiment is given below.

5.1.2.1 Independent variables

The four conditions are listed in order of their expected contribution to perceived quality of communication:

1. *Audio-only:* There was an audio connection and only a flat gray image on the video-tunnel monitor. The monitor was left on, both to provide greater consistency with the other three conditions, and to obscure the camera hidden behind the half-silvered mirror.

2. *Random-gaze avatar:* A head-and-shoulders view of an avatar appeared on the screen; depending on the gender of the participants taking part in each particular session, the avatar was either male or female. A decision was made to avoid using a fixed-gaze avatar since there is evidence [ALC74] that continuous gaze can result in negative evaluation of a conversation partner. It was therefore decided to use an avatar whose head and eyes moved, but in a way that was not related to the conversation except by accident.

3. *Inferred-gaze avatar:* The avatar was visually identical to the one in the *random-gaze* condition, but its head and eye movement received a different treatment. The avatar's head movement was determined by tracking of the participant's head, and the eye movement was determined by inference from conversational turn taking, as described in Section 5.1.2.5.

4. *Video:* The video-tunnel monitor showed a head-and-shoulders view of the remote participant with no perceptible audio or video delay. High-quality video was chosen as a baseline condition because in face-to-face interaction, factors such as interpersonal distance, posture and gesture might have confounded results [Arg88]. The purpose was to compare the performance of the avatar conditions against an optimal, but mediated, interaction.

5.1.2.2 Piloting

Section 3.1.4 stressed the importance of piloting. Particular emphasis was placed in this experiment on piloting the task. Since the two participants were expected to speak for several minutes and did not know each other prior to the experiment, it was necessary to give them a topic of conversation. Ten pilot sessions were conducted in order to test the feasibility of using a moral dilemma task in which participants were randomly assigned a point of view to argue from. The problem participants encountered, however, was to argue from a point of view that they did not necessarily agree with, on a topic that appeared to have a single ethically 'correct' standpoint. Conversations were stilted, participants quickly ran out of things to say, and they soon found themselves repeating the same points because they felt they had limited information about the situation. A new task addressing these drawbacks was assessed in additional pretests. The final task used for the experiment is described in Section 5.1.2.6.

Piloting also affected another central decision, namely whether or not to use a confederate for the conversation. For some of the pilot sessions, an experimenter acted as a confederate for the conversation task. While this presented the advantage of requiring half the number of participants, it raised concerns about habituation and fatigue effects. The four to six sessions expected per day, conducted over a period of weeks, meant there was no guarantee that the confederate's performance would be constant throughout. It was therefore decided to recruit pairs of participants instead.

5.1.2.3 Apparatus

The lab consisted of a reception room, a central control area from which the experiment could be monitored, and two separate soundproofed rooms in which the participants sat. In the control

area, the two experimenters could follow the conversation through headphones and could see a face-on view of both participants on the video wall, as illustrated in Figure 5.1. They could also use a microphone to speak simultaneously to both rooms, in order to inform participants when the time was up.

The soundproofed rooms, or 'cabinets', were equipped with identical equipment as described below. The rooms were purposefully bare in order to avoid providing visual distractions during the conversation.

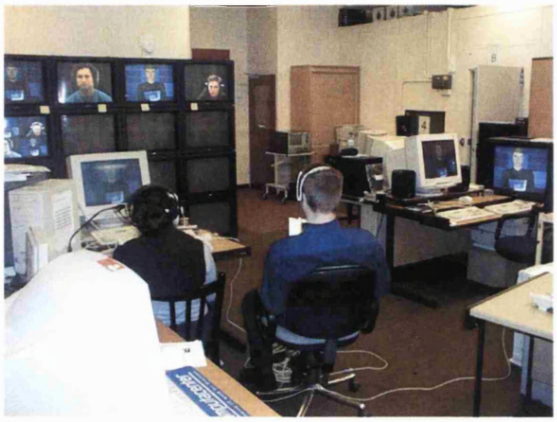


Figure 5.1: Both experimenters were able to follow the conversation on earphones while monitoring the participants and their corresponding avatars on the video wall.

The two cabinets in which participants sat were audiovisually linked using a video-tunnel. In each room a camera was placed behind a half-silvered mirror reflecting the image from an upturned monitor, as illustrated in Figure 5.2. Thus, while the participant looked at the reflected image, the hidden camera captured a face-on view that was sent to the remote participant's monitor. Participants sat two metres from a 21-inch Sony PVM-2130QM video monitor. The use of the video-tunnel setup had the added advantage of enabling a direct comparison between a video image and an equivalent head-and-shoulders view of an avatar. It also kept a fixed distance between the participants, ensuring that they were close enough to be able to discern the avatar's appearance and movement in sufficient detail.

Visual image: In the video condition each monitor was directly linked to the camera output from the other room. In the avatar conditions an S-video link with Vinegen Pro scan rate converters was used to connect each computer with the other person's monitor, enabling participants to see each other's avatar. Participant A's avatar was run on a Dell Dimension XPST 550 (Pentium III), Windows 98, with a GeForce 256 chipset, Gulillemot 3D Prophet video card and Creative AWE32 sound card. Participant B's avatar was run on a Compaq AP400 PIII 500,

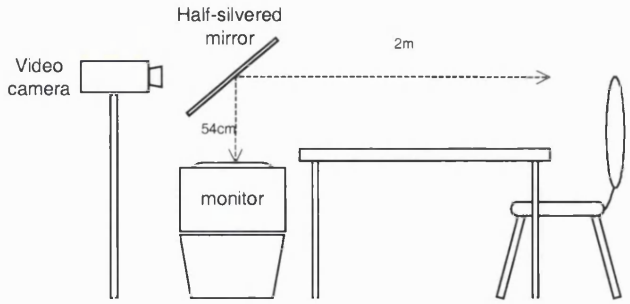


Figure 5.2: Video-tunnel setup

Windows 98, with a GeForce 256 chipset, Elsa Gloria2 video card and integrated sound.

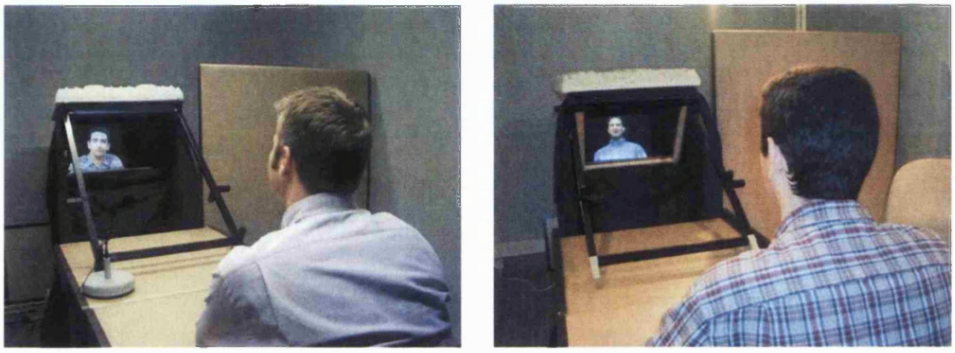


Figure 5.3: Video-tunnel setup showing the video condition

Audio: In all conditions, the sound was recorded using an AKG C747 microphone placed on the desk. In the video condition, participants heard their partner's voice through NXT flat panel speakers placed behind the screen; these are designed to limit feedback to the microphone, thereby minimising echo. However, a modification to the audio setup was necessary in the remaining conditions. Since the audio stream drove the avatar's lip and eye movements, it was necessary to isolate each microphone from incoming sound from the other room. Since the priority was to compare the avatar conditions with the audio-only condition in order to determine the contribution made by the avatar, it was decided that the same audio setup should be used in the audio-only condition. Participants were therefore equipped with Senneheiser HD265 headphones. A copy of the audio stream was sent to an audio matrix, where it was split into two, the first copy being sent to the conversation partner's headphones and the second to the computer to drive the avatar's behaviour. In the *inferred-gaze* condition, this audio information sent to the computer was accompanied by head position information from the Polhemus Isotrak II tracking system (Figure 5.4) .

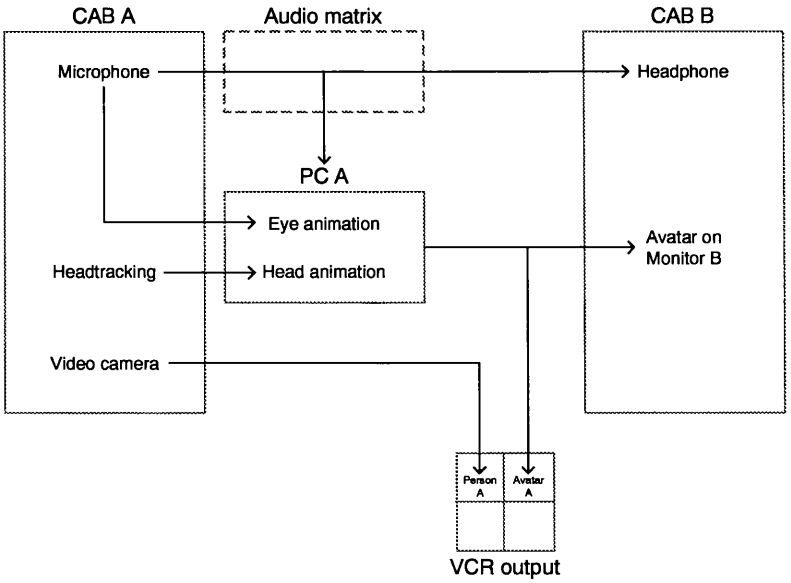


Figure 5.4: Diagram illustrating the video and audio output from one of the soundproofed rooms (CAB A) to the other room (CAB B) as well as to the videotape

5.1.2.4 Avatars

A female avatar was used for female participants, and a male avatar for male participants. Both participants in each pair were represented by a visually identical avatar as differences in facial geometry and texture mapping could potentially impact on the visual effect of the animations. However, the participants never saw their own avatar, so the fact that both were identical was unknown to them. Each avatar was independently controlled for each user.

The avatars were created by Norwich-based company Televirtual [Tel03] for the BT Talkzone exhibit in the Millennium Dome in London. They were originally designed to greet visitors entering the exhibit. The avatars were capable of a selection of behaviours such as smiling, frowning, looking sad, shrugging, pointing and waving controlled using keyboard presses. These behaviours could be triggered individually or in combination with each other. By default, each avatar's head made slight head and eye movements, based on rules described below.

5.1.2.5 Avatar animations

The animations originally designed for the BT Talkzone avatars were used, unaltered, for the *random-gaze* condition in this experiment. The timings and directions for both head and eye movement were determined using the computer's pseudo random number generator function and therefore had no relationship with the ongoing interaction. These original animations were ideal for the purposes of this experiment because they were designed to appear natural but were

in no way tied to the content or flow of conversation.

The perception of eye gaze depends on a combination of head and eye orientation [Arg88, AC76, GP63]. The head movements for the inferred-gaze avatar were tracked using a single Polhemus sensor attached to the headphones as illustrated in Figure 5.5.

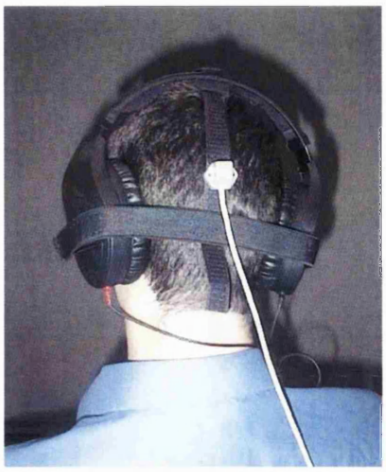


Figure 5.5: The sensor was attached to the headphones.

The eye movements were inferred from the audio stream. One of the fundamental rules for gaze behaviour in face-to-face dyadic interaction is that people gaze at their communication partner more while listening than while speaking [AIAM73, AC76, AI72]. Drawing on this principle, *while speaking* and *while listening* eye animations were implemented based on timing and frequency information taken from face-to-face dyadic studies [AIAM73, AC76, AI72]. The rules for these animations are summarised in Table 5.1.

Table 5.1: Summary of rules for 'while speaking' and 'while listening' animations

	While speaking	**While listening**
Begins with	'away' look	'at partner' look
Mean duration of 'at' gaze	1.8 seconds	2.5 seconds
Mean duration of 'away' gaze	2.1 seconds	1.6 seconds
Approximate frequency of glances	14/min	17/min

For 'at partner' gaze, the avatar's eyes focused directly ahead. The values for vertical and horizontal angles of 'away' gaze were chosen randomly from a uniformly distributed range of 0 to 15 degrees. The sign of the angle was random. In order to avoid repeating identical animation loops the duration of 'at' and 'away' gaze was randomised using the waiting time exponential distribution. Figures 5.6 and 5.7 show the male and female avatar looking 'at' and 'away'.

Figure 5.6: Face-on view of male avatar looking 'at' and 'away' from partner

Figure 5.7: Face-on view of female avatar looking 'at' and 'away' from partner

5.1.2.6 Task

In order to assess the avatar's impact on perceived quality of communication, a task was needed in which participants would be sensitive to visual feedback. It has been suggested [Sel95, SM94] that users benefit most from having visual feedback when performing equivocal tasks that have no single 'correct' outcome but require negotiation. Short, Williams and Christie argue that tasks involving conflict and negotiation are particularly suited to testing perceptions of communications media:

> "The negotiation situation is one which might be expected to be particularly sensitive to variation in the medium of communication. Each side's actions are highly contingent on his perceptions of the other's previous moves. Any changes in the communications link which might affect the information available about the other and thus distort interpretations of his actions could have strong implications for the eventual outcome. In view of this expected sensitivity to medium, the negotiation situation represents an important testing ground for theories about the origins of any effects of medium of communication" [SWC76, p.90].

A negotiation task was developed specifically for the study through a series of pilots, as detailed in Section 5.1.2.2. Participant A was asked to represent a mayor and participant B a baker. For female pairs both the characters were described as female, whereas for male pairs they were described as male. In the fictional scenario the baker's daughter was pregnant,

allegedly by the mayor's son. The son had refused to take responsibility, prompting the baker's irate spouse to draft a letter to the press. It was in the interests of both parties to avoid a scandal. Each participant was given slightly different goals and the task was to reach a mutually acceptable decision within ten minutes in order to prevent the letter from being mailed to the press (see Appendix D for the full scenarios).

5.1.2.7 Population

This experiment was conducted in the subjective testing department at BT Exact laboratories. For security reasons it was not possible to recruit from outside the laboratory, therefore participants were either permanent or temporary employees. Whether they were research or administrative personnel, as employees of the research division of a Telecommunications company they may have had a higher likelihood of exposure to novel communication technologies. For example, as many as 50% had used some form of video-mediated communication before, and 20% of those in the two avatar conditions had had prior exposure to virtual humans in some form.

Special care was taken to ensure that participants did not know each other. Short, Williams and Christie argue that "one might anticipate that media effects would be particularly marked when the interactors are relatively unacquainted. While people are still getting to know one another, any small additional piece of information can markedly affect overall judgments; later on in the acquaintance process, small changes in the available information would be expected to have less effect" [SWC76, p.112]. The negotiation task, combined with the fact that participants were unacquainted, meant that high demands were likely to be placed on the avatar.

Participants were recruited from a database of over 800 volunteers willing to participate in in-house experiments. The original goal was to have a minimum of 12 pairs of participants per condition. Extra sessions were run as a precautionary measure. Sessions with any irregularities were discarded, leaving a total of 50 pairs in the sample. Table 5.2 illustrates the factorial design.

Table 5.2: Factorial design for the first experiment on eye gaze

Condition	Male pairs	Female pairs	Total pairs
Video	11	1	12
Random-gaze avatar	10	2	12
Inferred-gaze avatar	11	2	13
Audio-only	10	3	13
Total pairs	42	8	50

It was difficult to recruit female participants owing to the demographics of the laboratory population. Nevertheless, an effort was made to recruit additional female volunteers through the 'women's network' mailing list in an attempt to increase the gender balance. All pairs were matched for gender and approximately by age, with the exception of a pair in the audio-only condition, where a man in his 30s was paired with one in his 40s.

5.1.2.8 Procedure

Participants were met in a reception area by two experimenters. Conversation between participants was avoided or kept to a minimum while they read the information sheet and signed the consent form (see Appendix C for the materials used). They were then led to two separate soundproofed rooms. Each participant was asked to sit down and the chair height was adjusted so that their face and shoulders were clearly visible on camera.

Participants were given ten minutes to prepare the scenario. Though they could take notes, they were informed that the piece of paper would be removed for the conversation. The video-tunnel monitors were already switched on when participants arrived. When both participants were ready the piece of paper describing the scenario was taken away to avoid visual distraction during the conversation. Participants were informed that the doors to the soundproofed rooms would be shut simultaneously and that the conversation would be timed from that point.

Conversations were stopped at the end of the allotted 10 minutes regardless of whether a conclusion had been reached. Participants were then asked to fill out a questionnaire. Each session concluded with a semi-structured interview, conducted individually with each participant. All of the conversations and interviews were videotaped with participants' written consent.

A slight change of procedure was required in the *inferred-gaze* condition, as it was necessary to calibrate the head tracking for each participant. First, participants were asked to sit straight while their chair height was adjusted and checked. Next, the experimenter manually calibrated the tracking system by observing an image of the participant's avatar on a separate monitor concealed behind the participant's chair.

5.1.2.9 Response variables

Perceived quality of communication was subdivided into four broad indicators: *face-to-face*, *involvement*, *copresence* and *partner evaluation*. The overall measure for each of these four responses was constructed from n questions, as detailed below.

1. **Face-to-face:** The extent to which the conversation was experienced as being like a real face-to-face conversation (n =6).

   (a) I could readily tell when my partner was concentrating on what I was saying.

(b) I was able to take control of the conversation when I wanted to.

(c) It was easy for me to contribute to the conversation.

(d) The conversation seemed highly interactive.

(e) There were frequent and inappropriate interruptions.

(f) This felt like a natural conversation.

2. **Involvement:** The extent to which the participants experienced involvement in the conversation (n =2).

   (a) I found it easy to keep track of the conversation.

   (b) I felt completely absorbed in the conversation.

3. **Co-presence:** The sense of being with and interacting with another person rather than with a computer interface (n =2).

   (a) I had a real impression of personal contact with my conversational partner.

   (b) I was very aware of my conversational partner.

4. **Partner Evaluation:** The extent to which the conversational subjects positively evaluated their partner, and the extent to which the conversation was enjoyed (n =5).

   (a) My partner was friendly.

   (b) My partner did *not* take a personal interest in me.

   (c) I trusted my partner.

   (d) I enjoyed talking to my partner.

   (e) I would be interested in meeting my partner face-to-face.

The responses to these variables were elicited by means of the post-experiment questionnaire, each response being on a 9-point Likert scale, where 1 was anchored to 'strongly disagree' and 9 to 'strongly agree'. The questions were partly based on previous questionnaires designed to elicit subjective responses to mediated communication [Sel95, SM94]. For the purposes of analysis the questionnaire anchors were swapped for negatively phrased questions such as 1(e) and 4(b) to ensure that 'high' scores would reflect a high score of the response variable being studied.

5.1.2.10 Explanatory variables

Data was collected on three auxiliary variables: gender, age, and whether or not the participant was a native English speaker. Additional questions were included concerning previous experience with videoconferencing and, for the avatar conditions, previous experience with virtual humans.

5.1.3 Results

The questionnaire data was analysed according to the method using the method of logistic regression described in Section 3.2.1. Figure 5.8 shows the means of the raw questionnaire responses. It is noteworthy that in every case, the video results in the highest score, and that in every case but for copresence the inferred-gaze avatar is second highest.

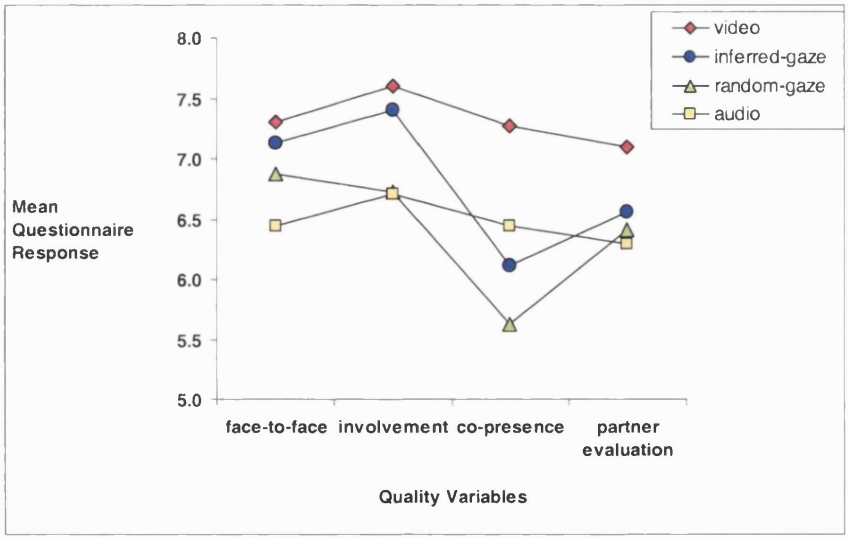


Figure 5.8: Means of raw questionnaire responses

Table 5.3 shows the mean and standard deviations of the counts of 'high' responses (7, 8 or 9) across the n questions in each condition.

Table 5.3: Mean $\pm$ standard deviations of count response variables

N = number of participants in each condition.

n = number of questions on which the count is based.

	Video (N=24)	Inferred gaze (N=26)	Random gaze (N=24)	Audio (N=26)
Face-to-face (n=6)	4.8 $\pm$ 1.4	4.5 $\pm$ 1.3	4.0 $\pm$ 1.6	3.4 $\pm$ 1.8
Involvement (n=2)	1.7 $\pm$ 0.4	1.6 $\pm$ 0.6	1.4 $\pm$ 0.7	1.3 $\pm$ 0.6
Co-presence (n=2)	1.6 $\pm$ 0.6	1.1 $\pm$ 0.8	0.7 $\pm$ 0.7	1.2 $\pm$ 0.7
Partner evaluation (n=5)	3.7 $\pm$ 1.4	3.2 $\pm$ 1.4	2.7 $\pm$ 1.5	2.8 $\pm$ 1.4

The results in Table 5.3 are consistent with the expectation that video will tend to produce the highest responses, and audio-only the worst (except for the copresence variable, where audio-only outperforms both avatar conditions). In each case the inferred-gaze avatar results in a higher count than the random-gaze avatar. Logistic regression is used to check for significance, and also to test whether other exogenous variables should be included in the model.

Table 5.4: Logistic regression with *face-to-face* as the response variable

Overall deviance = 214, d.f. = 95

The deviance column shows the increase in deviance that would result if the corresponding variable were deleted from the model.

Variable	Deviance χ^2	d.f.	5% χ^2
Condition	23.8	3	7.815
Role (baker)	8.4	1	3.841

Table 5.4 shows the results of the regression analysis for *face-to-face* as the response variable. Both condition and role were significant. In terms of *condition*, video and the inferred-gaze avatar were significantly different from the random-gaze avatar and audio-only conditions, but not significantly different from each other. *Role* refers to whether the participant played the mayor or the baker. The person who played the role of the baker tended to have a lower response count than the person who played the mayor.

Table 5.5: Logistic regression with *involvement* as the response variable

Overall deviance = 116.75, d.f. = 96

Variable	Deviance χ^2	d.f.	5% χ^2
Condition	8.9	3	7.815

A similar analysis was carried out for each of involvement and copresence as the response variables (Tables 5.5 and 5.6). The main-effect for condition was significant for both. For *involvement*, the inferred-gaze avatar is not significantly different from the video condition. The random-gaze avatar is significantly less than the video condition, but not significantly different from the audio. For *copresence*, video is significantly higher than all the others, and the worst case is the random-gaze-avatar. Role was again significant, with participants playing the baker tending to report a lower copresence response.

Table 5.7 shows the results of the regression for *partner-evaluation*. In this case both role and whether or not the person was a native English speaker are significant. Video again produces a significantly higher count than each of the other three conditions. The audio and

Table 5.6: Logistic regression with *copresence* as the response variable

Overall deviance = 147.8, d.f. = 95

Variable	Deviance χ^2	d.f.	5% χ^2
Condition	20.0	3	7.82
Role (baker)	4.47	1	3.84

Table 5.7: Logistic regression with *partner evaluation* as the response variable

Overall deviance = 182, d.f. = 94

Variable	Deviance χ^2	d.f.	5% χ^2
Condition	13.77	3	7.815
Role (baker)	12.22	1	3.841
Native Speaker	5.9	1	3.841

random-gaze avatar are not significantly different from each other, and the inferred-gaze avatar count is significantly higher than the random-gaze avatar and audio-only. The baker role again results in a lower count than the mayor role. Non-native English speakers tended to have a higher count than native speakers.

Table 5.8: Logistic regression for the sum of all response variables

Overall deviance = 329, d.f. = 95

Variable	Deviance χ^2	d.f.	5% χ^2
Condition	53.4	3	7.815
Role (baker)	28.5	1	3.841

Table 5.8 shows the results when all these response variables are combined. Once again, condition is highly significant, with video resulting in the highest count. The inferred-gaze avatar is significantly lower than video. The random-gaze avatar and audio-only conditions are again significantly lower than the others, but not different from each other. Also, the baker role has a lower overall count than the mayor role.

5.1.4 Experiment summary

This experiment deliberately isolated gaze animations by providing only a face-on, head-and-shoulders view of the avatar. The dual aim was to answer two questions. Firstly, whether an avatar can improve the perceived quality of communication between remote users. Secondly, whether an avatar whose gaze behaviour is directly related to the conversation can offer a significant improvement over a visually identical avatar with random gaze.

The random-gaze avatar did not provide a significant improvement over pure audio, suggesting that the simple introduction of an avatar does not automatically improve participants' perception of communication. Rather, the avatar must have certain behaviour characteristics in order to be useful. The inferred-gaze avatar outperformed the pure audio stream on several measures. This suggests that an avatar whose behaviours reflect an aspect of the conversation can indeed make a contribution to improving remote communication. Finally, the inferred-gaze avatar significantly outperforms the random-gaze avatar on all measures, indicating that an avatar whose behaviours are related to the conversation can present a marked improvement over an avatar that merely exhibits liveliness.

These findings had encouraging implications for inexpensive approaches to improving avatar fidelity. However, a central question remained unanswered. In the *inferred-gaze* condition, the avatar's gaze behaviour was being driven by two separate channels of information: its eye movement was based on inference from the audio stream, while its head movement was based on tracking the participant's real head movement. The open question was whether the significant impact was due to tracked motion data or from inferences about the eye movement based on research from face-to-face interaction. Answering this question would have significant implications for providing inexpensive ways to improve eye gaze based on information readily available from the audio stream.

Moreover, the experiment was conducted in a non-immersive setting. The question remained of how the inferred-gaze model would perform in a more demanding immersive setting, where participants were free to wander about a shared 3D space. The experiment presented in the next section was designed to address these concerns.

5.2 Experiment on Eye Gaze and Photorealism

The experiment presented in this section was designed to explore the relative impact of two logically distinct aspects of avatar fidelity: appearance and behaviour. In terms of behaviour, it directly extended the research presented in Section 5.1, and therefore shared several design similarities with the previous study. However, in addition to the impact of eye gaze, this experiment also investigated the effect of varying the photorealism of the avatar. Table 5.9 summarises the key distinctions between the two experiments.

Between the publication of results from the first eye gaze experiment [GSBS01] and the running of this second experiment, Lee, Badler and Badler published a similar study comparing subjective responses to a humanoid agent with static, random and inferred gaze [LBB02]. Their agents' inferred-gaze animations were consistent with the timings from the face-to-face litera-

Table 5.9: Summary of key differences between the two experiments on eye gaze

Experiment	**Eye gaze** (Section 5.1)	**Eye gaze + Photorealism** (Section 5.2)
Head-tracking and eye animations	Confounded	Disambiguated
Interface	Non-immersive (videotunnel)	Immersive (Cave/HMD)
Participant's position	Fixed relative to screen	Determined by participant's navigation in the VE
View of avatar	Head and shoulders	Full body
Avatar appearance	One level: semi-photorealistic	Two levels: visually simplistic and semi-photorealistic

ture detailed in Section 5.1.2.5, but were refined using a statistical model developed from their own gaze tracking analysis of real people engaged in dyadic interaction. Their results from a 12-person evaluation are consistent with the findings of the experiment reported in Section 5.1, in that the inferred-gaze model results in more positive perceptions. The inferred-gaze agent significantly outperforms the visually identical random-gaze agent in terms of perceived interest, engagement, friendliness, and liveliness. However, it is not clear whether participants were engaged in two-way verbal communication with the agent, or whether they simply viewed the animations on a screen.

In terms of eye gaze and photorealism, two studies by Fukayama et al. are also directly relevant [FSO+01, FTM+02]. The first is a 13-person study concerning the impact of eye animations on the impressions participants formed of an interface agent [FTM+02]. Their gaze model consists of three parameters: amount of gaze, mean duration of gaze and gaze points while averted. Their comparative analysis of responses to nine different gaze patterns suggests that agent gaze can reliably influence impression formation. For this particular study they isolated the agent's eyes from any other facial geometry. In a related study, they investigate whether the impact of the gaze patterns is affected by the photorealism of the agent's face [FSO+01]. Their findings suggest that varying the appearance from visually simplistic to more realistic has no effect on the impressions produced. The interaction is one-way, with participants viewing a pre-recorded agent animation. It is therefore difficult to know whether the findings would generalise to a sustained verbal interaction.

One aspect of studies to date is that participants were shown a limited, head-and-shoulders view of the virtual human, and that the spatial relationship was fixed by the 2D nature of the interaction. They leave open the question of how these gaze models might hold up in an immersive situation where participants are able to wander freely around a shared space, and where they can interact with a full-body avatar.

5.2.1 Experimental aims and expectations

The goal for this experiment was threefold. Firstly, to disambiguate between the effect of inferred eye movements and head-tracking, both of which may have contributed to the results reported in Section 5.1. Secondly, to test how the inferred-gaze model performs in a less forgiving immersive setting where it is not desirable to attempt to control the participant's gaze direction. Finally, to explore the combined impact on quality of communication of eye gaze model and visual appearance. The initial hypothesis was that behavioural realism would be independent in its effects on quality of communication from the impact of visual realism, and that the behavioural realism would be of greater importance. The inferred-gaze model was expected to outperform the random-gaze one for both the higher-realism and lower-realism avatar. One open question concerned the extent to which the gaze animations would impact on the lower-realism avatars, or how the two avatars would perform in comparison with each other.

5.2.2 Experimental design

5.2.2.1 Independent variables

This experiment investigated two distinct aspects of avatar fidelity: appearance and behaviour. A between-groups, two-by-two factor design was employed with the two factors being the avatar's level of photorealism, and its behavioural realism specifically in terms of eye gaze.

5.2.2.2 Piloting

Piloting took place over a period of one month. The task used was the same as in the previous experiment, so on this occasion the bulk of the pilot sessions concerned the experimental procedure. This was complex and required careful coordination between two experimenters in order to fit each session into the one-hour time limit.

Feedback from pilot participants emphasised the need to create an appropriate environment. Based on the grounded theory analysis of the interviews from the previous experiment (see Chapter 6), it was decided that participants should only see each other's avatar at the start of the conversation. Slater et al.'s experiment on virtual rehearsal [SHS+00] used a layout consisting of two large 'waiting' areas linked by a corridor to the kitchen area where the actors were to rehearse their scene. Although this virtual model did not lend itself to the present scenario, it served as inspiration for a purpose-built environment consisting of two training rooms linked to a central meeting room. This model ensured that when the connecting doors were opened, participants would see their partner's avatar directly in front of them as they entered the central room (see Figure 5.12).

While the experiment on responsiveness (Chapter 4) involved an active exploration of the

VE, the conversation task in this experiment did not place particular emphasis on navigating around the space. However, the pilot sessions identified the need for a brief training period that had not been foreseen in the original design of the experiment.

5.2.2.3 Apparatus

Head-mounted Display (HMD): The scenarios were implemented on a Silicon Graphics Onyx with twin 196 MHz R10000, Infinite Reality Graphics and 192M main memory. The tracking system has two Polhemus Fastraks, one for the HMD and another for a 5-button 3D mouse. The helmet was a Virtual Research V8 which has true VGA resolution with 640x480x3 colour elements for each eye. The V8 has a field of view of 60 degrees diagonal at 100% overlap. The frame rate was kept constant for both the Cave and the HMD.

ReaCTor: The Cave-like system used was a ReaCTor made by Trimension, as described in Section 4.2.3.

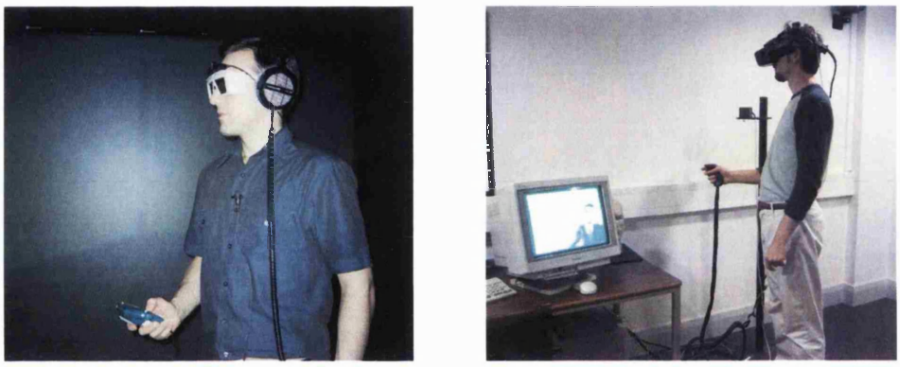


Figure 5.9: Participants in the Cave (left) and HMD (right). Participants in the HMD could not see their own bodies or physical surroundings while in the IVE; the image visible on the monitor was for the benefit of the experimenters.

Both participants had wireless microphones attached to their clothing; these were activated only for the duration of the conversation. Figure 5.9 shows one participant in the Cave and the other in the HMD.

5.2.2.4 Avatars

Participants were represented to their conversation partner as a life-size avatar, as illustrated in Figure 5.10. Both participants in each pair were represented by a visually identical avatar to avoid differences in facial geometry affecting the impact of the animations. Participants were unaware of this because they did not see their own avatar in full. The participants in the HMD, who were visually isolated from the physical surroundings of the lab, could see the hands and feet of their avatar when looking down; the participants in the Cave could only see their own

physical body.

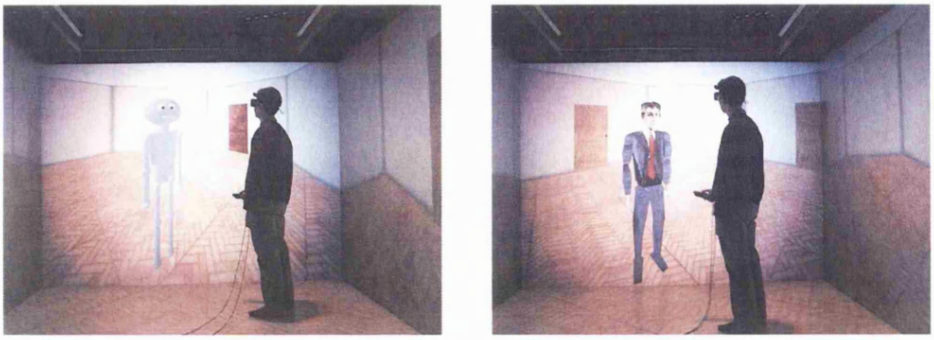


Figure 5.10: Participants saw their conversation partner as a life-size avatar. The avatar was either lower-realism (left) or higher-realism (right)

Though identical in appearance, the two avatars were of course animated separately for each user. Again, the software for the VE was written on top of the DIVE software platform [FSS+01] (see Section 4.2.3.1). A plugin was developed in C to animate the avatar body parts. DIVE reads the user's input devices and maps physical actions to logical actions in the DIVE system. In this case the head and the right hand were tracked. At the start of each session, the avatars were moved to their correct starting positions in their corresponding training rooms with the aid of Tcl script. A separate Tcl script was used to open the doors separating the virtual rooms at the end of the training period.

In the lower-realism condition a single, genderless avatar was used to represent both males and females. For the higher-realism avatar, a male and female avatar were used, as shown in Figure 5.11.

Figure 5.11: From left to right: Lower-realism avatar, higher-realism female avatar, higher-realism male avatar

All avatars used in the experiment were made H-Anim compliant [Hum03] and had identical functionality. A plugin was used to animate the avatar's body in order to maintain a visually consistent humanoid. This included inferring the position of the right elbow using inverse kinematics when the user's tracked hand moved, and deducing the position of the avatar's knees

when the user bent down. There were also some deductions involved in the rotation of the head and body. The body was not rotated to the same direction as the head unless there was some translation associated with the user. This was to enable the user to nod, tilt and shake their head in the VE whilst in conversation.

One of the central aims of this experiment was to disambiguate the impact of head-tracking and the inferred eye animations. Therefore participants' heads were tracked in all conditions, and only the eye animations were varied as illustrated in Table 5.10. The random-gaze eye animations were identical to the first eye gaze experiment. The inferred-gaze animations were refined based on newly published information.

Table 5.10: The second experiment on gaze used head-tracking in all conditions, only varying the eye animations

Eye gaze experiment (Section 5.1)

	Random-gaze	**Inferred-gaze**
Head	Random	Tracked
Eyes	Random	Inferred

Eye gaze + photorealism experiment (Section 5.2)

	Random-gaze	**Inferred-gaze**
Head	Tracked	Tracked
Eyes	Random	Inferred

After the publication of findings from the first eye gaze experiment [GSBS01], Lee, Badler and Badler [LBB02] published a closely related study comparing responses to a humanoid agent with random and inferred gaze. They based their animations on their own empirical gaze tracking research. Their model is consistent with timing expectations from the literature, but adds valuable new probabilities for gaze direction during 'away' fixations that were absent in the first experiment. Both models were implemented for the *random-gaze* condition and were compared in a pre-test. Although they did not appear to be visually distinguishable, Lee et al.'s model was selected since it was based on more detailed eye gaze data.

Both previous models assumed a non-immersive setting where the participant was seated in front of a screen. The avatar's 'at partner' gaze was therefore always fixed straight ahead. In this immersive experiment, a decision was made not to automatically target 'at partner' gaze at the other avatar. Rather, 'at partner' gaze was kept consistent with the position and orientation of the head. In this way, the avatar could only seem as if it was looking 'at partner' if the participant was in fact looking directly at the other avatar's face (based on head-tracking information).

5.2.2.5 Virtual environment

The shared IVE in which the participants met consisted of two spacious 'training' rooms connected to a smaller 'meeting' room in the centre, as illustrated in Figure 5.12. The doors separating the virtual rooms were kept closed during the training session to avoid participants seeing each other's avatar before the conversation task. All rooms were kept purposefully bare so as to minimise visual distraction.

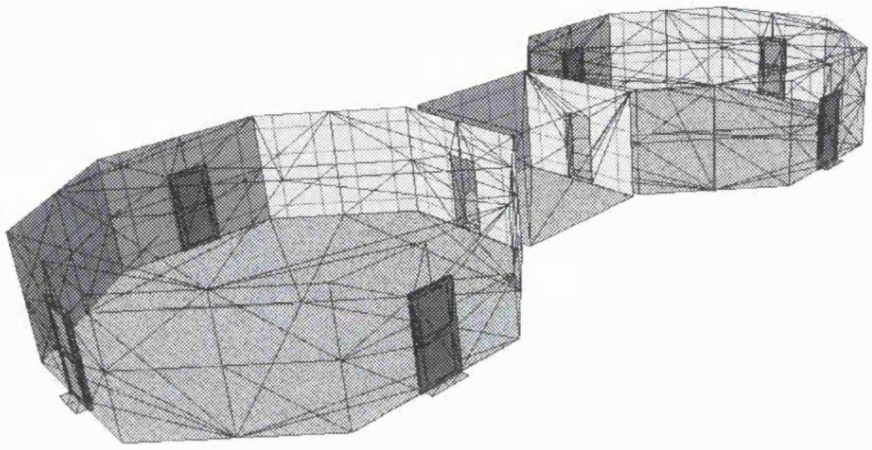


Figure 5.12: Two separate training rooms were separated from a central meeting room by doors that were only opened at the start of the conversation task.

5.2.2.6 Task

The same role-playing negotiation task was used as in the previous eye gaze experiment. The only addition was a diagram illustrating the relationships between the characters in the scenario, to help participants remember the necessary information (Appendix D).

5.2.2.7 Population

48 participants were recruited from the UCL campus using an advertising poster campaign. The poster informed prospective participants that they were required to have good vision (eyeglasses were acceptable), be proficient speakers of English, and not be subject to epileptic seizures. Participants were paid £5 for the one-hour study. They were paired with someone of their own gender and assigned randomly to one of the four conditions. They did not know their conversation partner prior to the experiment, and were not allowed to meet beforehand. A gender balance was maintained across the four conditions, as illustrated in Table 5.11. The reason for this is that there is evidence that males and females can respond differently to nonverbal

behaviours, particularly in the case of eye gaze cues [AI72].

Table 5.11: Factorial design for the experiment on eye gaze and photorealism

	Random gaze	Inferred gaze
Lower-realism avatar	3 male pairs 3 female pairs	3 male pairs 3 female pairs
Higher-realism avatar	3 male pairs 3 female pairs	3 male pairs 3 female pairs

A total of 7 pairs were discarded because technical difficulties had created irregularities in the procedure. Irregularities included the audio connection not working correctly, and on one occasion the Cave working in mono as opposed to stereo, so that the image was not experienced in 3D. The remaining 24 pairs were kept in the sample. The age of participants ranged from 10 to 52. Table 5.12 and Figure 5.13 detail the age distribution for both male and female participants.

Table 5.12: Age distribution for participants

Age	under 21	21-25	26-30	31-35	36-40	41-45	over 50
Male	2	9	6	3	2	2	0
Female	0	8	12	1	2	0	1
Total	2	17	18	4	4	2	1

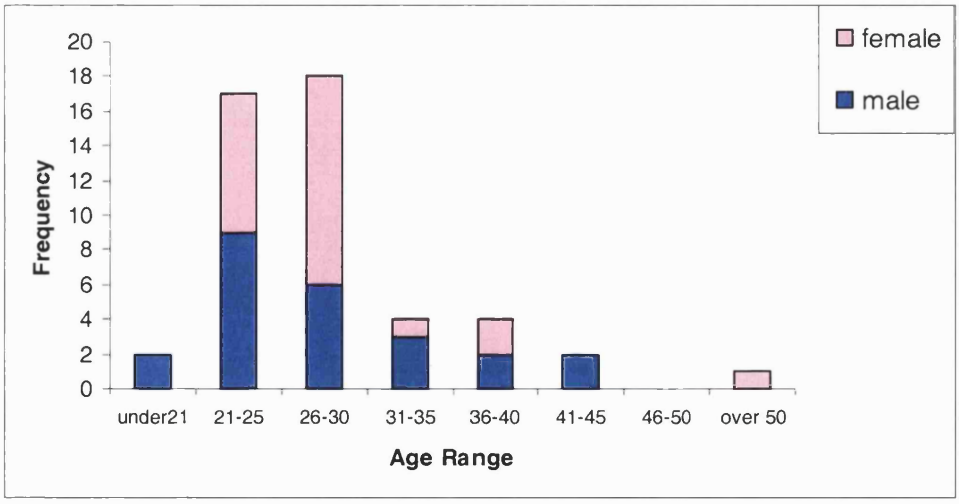


Figure 5.13: Age range for participants

5.2.2.8 Procedure

In numerous ways the design of this experiment was facilitated by lessons learned through the previous eye gaze experiment. The grounded theory analysis to be reported in Chapter 6 revealed that even a brief glimpse of the conversation partner could significantly affect responses to the avatar. Participants therefore did not meet prior to the experiment, to avoid the possibility of any first impressions influencing the role of the avatar in the conversation. The first person to arrive was assigned to the Cave, the second to the HMD in an adjacent room. The assigned interface (Cave or HMD) was counterbalanced with role (mayor or baker) to avoid introducing bias. Participants filled out a computerised background questionnaire and then read the scenario (see Appendix D for the materials used in this experiment).

The apparatus was prepared and tested prior to each session in order to ensure the correct placement of the avatars in their respective training rooms. Participants were given five to ten minutes to perform a navigation task in their separate training rooms. Once they felt comfortable with navigation, both participants were asked to face the door leading to the central meeting room. They were told that shortly the doors would open onto an adjacent room where they would be meeting their conversation partner, and that the conversation could begin at that point. The doors separating the virtual training rooms from the central meeting room were opened simultaneously and the microphones were activated at the same point so they could hear each other.

During the conversation, both experimenters observed their participant and filled out an observation sheet that covered a number of items including left or right handedness, amount of movement around the virtual space, amount of head nods and head shakes, gesticulation, and any additional comments. Conversations were stopped after 10 minutes regardless of whether an agreement had been reached. At the end of the conversation the audio connection between the participants was disconnected to avoid them hearing each other's comments immediately after the experience. The session concluded with a post-questionnaire and a semi-structured interview conducted individually with each participant. Finally, participants who wished to do so were allowed to meet their conversation partner in person.

5.2.2.9 Response variables

This experiment was designed to directly extend the previous experiment on eye gaze. Consequently, the same four indicators of perceived communication quality of communication were considered in the analysis: *face-to-face*, *involvement*, *copresence* and *partner evaluation*. The questionnaire items for each of these variables were detailed in Section 5.1.2.9.

A number of additional variables were considered in the analysis, including *gaze realism*,

perceived *avatar fidelity*, *presence* and a more detailed measure of copresence called *social-copresence*. Each of these responses consisted of *n* questionnaire items, as described below.

Gaze fidelity (n=2): The degree to which participants experienced a sense of gaze with their partner.

1. My partner and I frequently made eye contact.
2. The way my partner looked at me appeared natural.

Avatar fidelity (n=4): The degree to which the avatar's appearance and behaviour were seen to be realistic.

1. The avatar's appearance was realistic.
2. The avatar's behaviour was realistic.
3. The avatar was *not* expressive.
4. The avatar did *not* always seem to respond appropriately to me.

Social-copresence

The copresence measure used in the first study on eye gaze consisted of two questionnaire items drawn from previously published studies [Sel95, SM94]. The notion of copresence became increasingly defined during the course of the research, partly as a result of the experiment on responsiveness (Chapter 4) and the grounded theory analysis (Chapter 6). This third and final experiment therefore considered a more detailed measure of copresence consisting of a number of subcomponents. Purely in the interest of clarity, this will be referred to here as *social-copresence*. This response variable consisted of the following subcomponents: *general copresence*, *spatial copresence* and *personal contact*.

1. **General copresence** (n=5): The extent to which participants felt they were in the company of another person. This variable consisted of 'generic' descriptors of copresence.
   (a) I felt alone.
   (b) I did not feel my partner and I were together.
   (c) I had a sense of being in the company of my conversation partner.
   (d) I had a sense of being with the other person.
   (e) I behaved as if there was nobody in the virtual room with me.

2. **Spatial copresence** (n=2): The extent to which participants felt they were in the same space as their partner.

   (a) I felt my partner and I were in a shared space.

   (b) It did *not* feel as though my partner and I were in the same room.

3. **Personal contact** (n=2): The degree to which participants experienced a sense of personal contact with their partner.

   (a) I had a real impression of personal contact with my conversational partner.

   (b) The interaction did not seem very personal to me.

Presence (n=5): Seeing as participants met immersively, they were also asked to report on their sense of presence. The 5 questionnaire items from the SUS questionnaire were used [SSU93], as detailed in Section 4.2.9.

Whilst the previous experiment on eye gaze used a 9-point Likert scale, each questionnaire response in this study was on a 7-point Likert-type scale where 1 was anchored to strong disagreement and 7 to strong agreement. This was in the interest of consistency with the SUS presence questionnaire items, that use a 1-7 scale [SSU93]. Again, some questionnaire anchors needed to be swapped for the analysis so that all 'high' scores would reflect a high score of the response variable being studied.

5.2.2.10 Explanatory variables

As well as the independent variables (two levels for both visual and behavioural fidelity), a number of explanatory variables were used in the analysis. These included gender, age, and status (status referred to participants' occupation, if part of the university). The grounded theory analysis for the first experiment (Chapter 6) suggested that participants' previous technical and media experience might shape their responses to interacting with virtual humans. The experiment on responsiveness (Chapter 4) therefore collected explanatory data on computer literacy. This final experiment did so with increased precision by investigating participants' technical expertise in terms of computer use and programming, as well as experience with interactive virtual reality systems and computer games (see the 'background information' questionnaire in Appendix D). Another important explanatory variable was the degree of participants' social anxiety in everyday life, measured again using the standard SAD questionnaire [WF69].

5.2.3 Results

This section presents the findings for the subjective response variables detailed above. Figures 5.14 and 5.15 show the means of raw questionnaire responses for the lower-realism and

higher-realism avatars. The questionnaire data was again analysed using the method of logistic regression described in Section 3.2.1.

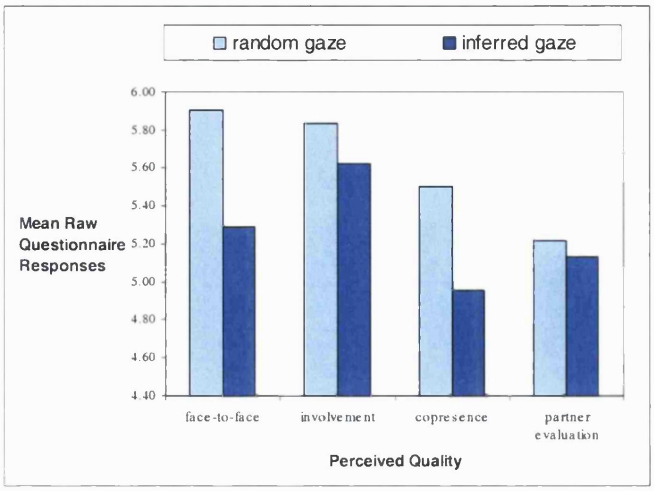


Figure 5.14: Means of raw questionnaire responses for lower-realism avatar

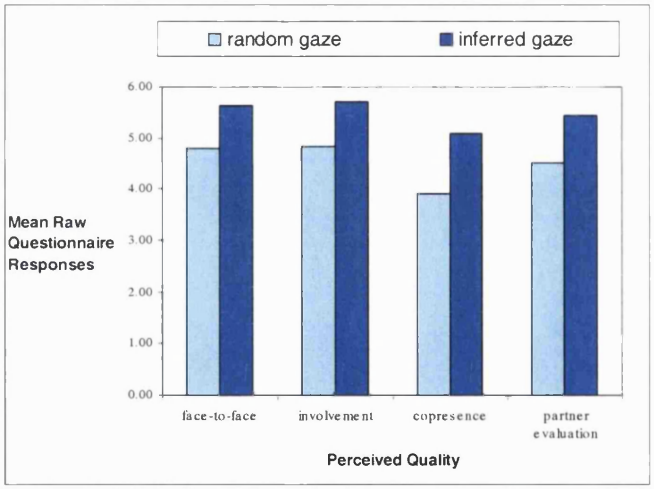


Figure 5.15: Means of raw questionnaire responses for higher-realism avatar

Table 5.13 shows the means and standard deviations of the counts of 'high' responses (6 or 7) across the n questions in each condition. An inspection of the *face-to-face* response suggests that there is a strong interaction effect: within each row and column there is a significant difference between the means, but the direction of the effect is different for random and inferred gaze.

Table 5.14 shows the the increase in deviance that would result for each response variable if the corresponding predictor variable were deleted from the model. Each of these terms is sig-

Table 5.13: Mean ± standard deviations of count response variables for perceived quality of communication

Response	Type of avatar	Random Gaze	Inferred Gaze
Face-to-face	Lower-realism	4.2 ± 0.5	2.9±0.5
	Higher-realism	2.2±0.4	3.9±0.6
Involvement	Lower-realism	1.3±2.9	1.3±0.2
	Higher-realism	0.9±0.2	1.2±0.2
Copresence	Lower-realism	1.2±0.2	0.7±0.2
	Higher-realism	0.3±0.1	1.1±0.3
Partner Evaluation	Lower-realism	2.6±0.5	2.2±0.4
	Higher-realism	1.8±0.5	2.8±0.5

Table 5.14: Logistic regression for response variables for perceived quality of communication

The tabulated χ^2 5% value is 3.841 on 1 d.f., and all d.f.s are 1.
The deviance column shows the increase in deviance that would result if the corresponding variable were deleted from the model.

	Deviance χ^2			
Fitted Variable	**Face-to-face**	**Involvement**	**Co-presence**	**Partner evaluation**
Type avatar • type gaze	22.03	-	9.7	5.0
Age	7.8 (+)	16.9 (+)	14.1 (+)	-
Role (baker)	10.0	-	-	6.2
SAD	15.7 (-)	-	-	-
Overall deviance	79.9	67.7	60.5	125.0
Overall d.f.	40	46	43	44

nificant at the 5% level of significance (i.e., none can be deleted without significantly reducing the overall fit of the model). *Type of avatar* and *type of gaze* were significant for 3 of these 4 response variables. The participant *age, role* and *SAD* score were significant for some responses (*role* refers, again, to whether they played the mayor or baker in the negotiation task). Just as in the previous gaze experiment, the person who played the role of the baker tended to have a lower response count than the person who played the mayor.

The results for *face-to-face* as the response variable illustrate the analysis. The type of *interface* (Cave or HMD) did not have a significant effect on responses. However, the *age* of participants was found to be significant, and positively associated with the response: older people were more likely to have rated their experience as being like a face-to-face interaction.

The formal analysis demonstrates the very strong interaction effect between the *type of avatar* and the *type of gaze* (denoted by the • symbol in Table 5.14). In other words the impact

of the gaze model is different depending on which type of avatar is used. For the higher-realism avatar the (more realistic) inferred-gaze behaviour increases *face-to-face* effectiveness. For the lower-realism avatar, the (more realistic) inferred-gaze behaviour reduces *face-to-face* effectiveness. This is illustrated by Figures 5.14 and 5.15 above, showing the means of raw questionnaire responses for each avatar.

For the lower-realism avatar, the inferred-gaze model has a consistently negative effect on each response variable (Figure 5.14). The opposite is true of the higher-realism avatar (Figure 5.15). This seems to indicate a need for consistency between the visual appearance of the avatar and the type of behaviour that it exhibits. With respect to eye gaze, low fidelity appearance demands low fidelity behaviour, and correspondingly higher fidelity appearance demands a more realistic behaviour model. The logistic regression analysis suggests that there is a significant interaction effect between *type of avatar* and *type of gaze* for 3 out of the 4 response variables. The exception is *involvement*, for which there is no significant effect of either avatar or gaze type.

Table 5.15: Logistic regression for *social-copresence* and its subcomponents (*general copresence*, *spatial copresence* and *personal contact*)

The tabulated χ^2 5% value is 3.841 on 1 d.f., and all d.f.s are 1.

	Deviance χ^2			
Fitted Variable	**Social-copresence**	**General copresence**	**Spatial copresence**	**Personal contact**
Type avatar • type gaze	21.03	4.06	17.42	13.99
Age	39.02 (+)	20.81 (+)	11.13 (+)	19.66 (+)
SAD	14.08 (-)	19.23 (-)	-	-
Programming	9.48 (-)	-	5.21 (-)	-
Overall deviance	144.51	131.72	67.47	47.47
Overall d.f.	40	41	42	43

Table 5.15 shows the results of the regression analysis for *social-copresence* and its subcomponents, *general copresence*, *spatial copresence* and *personal contact*. The analysis shows the same strong interaction effect between the *type of avatar* and *type of gaze* for the *social-copresence* response, as well as for each of its subcomponents. *Age* is also significant for each variable, with older participants more likely to report a higher response. The *SAD* score had a significant negative effect on *social-copresence* as well as the *general copresence* subcomponent. In addition, programming experience was found to be significant, and negatively associated with both *social-copresence* and *spatial copresence* responses.

Table 5.16 shows the results of the logistic regression for *gaze fidelity* and *avatar fidelity*.

Table 5.16: Logistic regression with *gaze fidelity* and *avatar fidelity* as the response variables

The tabulated χ^2 5% value is 3.841 on 1 d.f., and all d.f.s are 1.

	Deviance χ^2	
Fitted Variable	**Avatar fidelity**	**Gaze fidelity**
Type avatar • type gaze	5.2	8.77
VR experience	4.24 (-)	-
Overall deviance	62.08	37.1
Overall d.f.	43	44

Previous experience with interactive VR systems was found to be significant, and negatively associated with perceived avatar fidelity. Overall the analysis indicates that there is an overwhelmingly cohesive model, where the same interaction effect between the *type of avatar* and the *type of gaze* holds for both perceived gaze fidelity and avatar fidelity, as well as for each subcomponent of both perceived quality of communication and social-copresence.

Table 5.17: Logistic regression with *presence* as the response variable

Overall deviance = 94.7, d.f. = 41

The tabulated χ^2 5% value is 3.841 on 1 d.f., and all d.f.s are 1.

Variable	**Deviance χ^2**
Age	43.2 (+)
SAD	7.7 (-)
Programming	8.9 (-)
VR experience	12.4 (-)
VR knowledge	14.5 (+)

Table 5.17 shows the results of the logistic regression for *presence*. Although neither *type of avatar* nor *type of gaze* significantly affected reported presence, a number of person-variables had a significant effect. The *age* of participants was again significant, with older participants more likely to report a higher sense of presence. The *SAD* score, level of *programming experience* and *VR experience* (previous experience of 'virtual reality') all had a significant effect, and were negatively correlated with presence. However, knowledge about how VR works was instead positively correlated with presence.

The results indicate a highly significant positive correlation between *presence* and *social-copresence* (Δdeviance=58.18, d.f.=1, p=0.000). This is consistent with the findings from the experiment on responsiveness (Section 4.3) suggesting a significant positive correlation between presence and copresence.

In addition to perceived quality of communication, other social responses were captured by the questionnaire. These included:

1. the extent to which the avatar was perceived as real and like a human;
2. the degree of association made between conversation partner and avatar;
3. the extent to which participants paid attention to the avatar during the conversation;
4. finally, the avatar's perceived contribution to helping participants to understand aspects of their partner's behaviour and attitude.

The findings for these additional variables are included in Appendix E.

5.2.4 Experiment summary

The study reported in Section 5.2 sought to investigate the impact of visual and behavioural fidelity in avatars on perceived quality of communication between participants meeting in a shared IVE. In terms of appearance, the avatar was either visually simplistic or more realistic. In terms of behaviour, eye gaze was singled out, comparing inferred-gaze and random-gaze models previously tested in a non-immersive setting. The findings clear up the ambiguity from the first experiment regarding whether the significant differences in performance between the gaze models were due to head-tracking or eye animations inferred from the audio stream. They indicate that independent of head-tracking, inferred eye animations can have a significant positive effect on participants' responses to an immersive interaction. The caveat is that there should be some consistency between visual and behavioural realism, since the lower-realism avatar did not appear to benefit from the higher-realism, inferred gaze model. This finding has implications for inexpensive ways of improving avatar expressiveness using information readily available in the audio stream. It suggests avenues for interim solutions for the difficult problem of providing robust eyetracking in a Cave-like systems.

5.3 Chapter Summary

This chapter presented two experiments designed to investigate the impact of varying levels of avatar fidelity on subjective responses to dyadic interaction with another person represented by an avatar. The experiment presented in Section 5.1 focused on the impact of eye gaze in the context of a non-immersive interaction. Findings suggested that the inferred-gaze avatar significantly outperformed the random-gaze avatar. However, there was a need to disambiguate between the impact of the eye animations inferred from the audio stream and real motion information captured by the head-tracker. The experiment reported in Section 5.2 addressed this

ambiguity by tracking head movement in all conditions and only altering the eye animations. Results suggest that the eye animations do indeed have an impact, independently of the head-tracking. Regarding the second question, the findings suggest that the inferred-gaze animation can outperform the lower-realism random-gaze animations, even in the more demanding immersive context. However, in response to the third and final question, it appears that the effect of the animations changes in relation to the avatar's appearance. Findings suggest a strong interaction effect between visual and behavioural fidelity, suggesting that there is a need to maintain a consistency between the two.

Chapters 4 and 5 have presented the findings from three experiments concerning responses to agents and avatars. Chapter 6 will present the in-depth qualitative analysis of the interview data from the first experiment on eye gaze. Chapter 7 will discuss the implications of the collective findings.

Chapter 6

Qualitative Analysis of Interview Data

As discussed in Section 3.2.3, the grounded theory method of qualitative analysis is specifically designed to allow the discovery of new themes from data, lending itself particularly well to the exploration of new phenomena about which relatively little is yet known. This chapter presents the grounded theory analysis of the interviews from the first experiment on eye gaze, expanding on the results from the questionnaire data reported in Section 5.1.3. The questionnaires focused on perceived communication quality, measured on a number of indicators: the extent of similarity to a face-to-face conversation, the degree of involvement and copresence felt by participants, and their impressions of their conversation partner. The findings indicated that the inferred-gaze avatar significantly outperformed both the random-gaze avatar and the audio-only conditions on three of the response measures, suggesting that an avatar can begin to contribute to mediated communication providing its behaviour reflects an aspect of the conversation taking place.

One of the reasons for conducting interviews is to obtain a richer description of participant responses than is possible through questionnaires. Some previous experiments on avatar-mediated interaction such as [SHS+00] have relied on anecdotal reporting of debriefing sessions to shed light on statistical findings. Some 'general themes' emerged from a first reading of the interview transcripts from the first eye gaze experiment. Although the grounded theory analysis that followed was consistent with these general themes, it provided a richer and more detailed explanation of how different aspects of avatar fidelity can shape subjective perceptions.

After describing the agenda for the semi-structured interviews and the general themes identified after an initial reading of the transcripts, this chapter will describe how Strauss and Corbin's method [SC98] was adapted in the analysis. It will then introduce the key categories that emerged in the analysis, followed by a narrative 'storyline' that links them together. It will conclude by focusing on the impact of visual and behavioural fidelity highlighted by the analysis.

6.0.1 Method of data collection

Semi-structured interviews were conducted individually with each participant in all four conditions: video, inferred-gaze avatar, random-gaze avatar and audio-only. Since the qualitative analysis focuses on the role of the avatar, only the 48 interviews with participants from the two avatar conditions were kept in the data pool.

6.0.1.1 Semi-structured interviews

Participants were interviewed individually by their experimental minder in the same room where they had carried out the conversation task and filled out the questionnaire. During the interviews the avatar was left on the video-tunnel screen, allowing participants to point out relevant individual features or behaviours relevant to the discussion. The session schedule generally restricted interviews to between 10 and 20 minutes, though participants wishing to speak for longer were allowed to do so wherever possible.

According to Smith [Smi95] the goal of semi-structured interviews is to cover areas of interest while remaining sufficiently flexible to allow the respondent to focus on issues of particular individual concern. It was often the case that participants expressed a greater interest in a specific aspect of the experience, such as the audio quality or the avatar's lip synch. In cases where the conversation became too focused on an issue that was only of peripheral interest, such as the negotiation strategy employed in the conversation task, participants were gently steered towards more relevant questions. The main questions covered in the semi-structured interview are summarised in Appendix C.

In keeping with the flexible approach of semi-structured interviews, questions were not posed in a fixed order. Often participants would spontaneously address two or more issues in detailed response to a single question. The semi-structured interviews expanded as the experiment went on to include questions highlighted by participants in previous sessions. For example, it became evident that some people had a mental image of the person they were speaking to that was separate from either the avatar or the real person they had briefly seen upon arrival in the laboratory. This seemed highly relevant and was covered in successive interviews.

The interviews were transcribed verbatim from audio-only copies of the video tapes. Interviews were labelled to identify the condition (C) and group (G) the participant was in, and whether they played the mayor (A) or baker (B). For example, C2G3A means that the participant played the mayor in the third group to take part in condition 2 (random-gaze avatar); condition 3 is the inferred-gaze avatar. A sample transcript is included in Appendix F.

6.1 Impressions from Reading the Interview Transcripts

Upon careful reading of the questionnaire transcripts, it was possible to anecdotally identify some emerging themes. There appeared to be three main threads to people's reactions: negative, neutral and positive. The main themes are summarised in Table 6.1.

Table 6.1: General themes identified prior to grounded theory analysis

Thread	Theme
Negative	• Avatar is a visual distraction.
Neutral	• Avatar is not distracting, but it fails to contribute to conversation. • Aural and visual information are disconnected.
Positive	• Avatar adds value by providing a visual focus. • Avatar adds interactivity as compared to a telephone call

Although the overall responses ranged from negative to neutral to positive, the incidence of positive statements was higher in the *inferred-gaze* condition, as consistent with the outcome of the qualitative analysis of the questionnaires. Several participants also commented that the amount and quality of the avatar's eye movement was inappropriate. They complained that it was too animated and that it looked away too much given the serious situation being discussed. As it was, the avatar didn't give them the impression of listening intently enough to the conversation.

Respondents complained repeatedly about two things. Firstly, the lack of facial expression, resulting in an almost complete absence of emotional information; secondly, the mismatch between visual cues and the cues available from the audio stream. Since participants believed that the voice was human whereas the avatar was synthetic, they placed far greater reliance on the voice. This section has identified some of the general themes that emerged from a general reading of the transcripts. The following section summarises the goals of the grounded theory analysis.

6.2 Goals of the Analysis

The general impressions from reading the interviews cluster around a selection of themes: insufficient feedback, the mostly negative impact of the eye gaze models, and the high reliance placed on the verbal exchange. Though anecdotal impressions are illustrative, they do not give predictive power. The goal of the grounded theory analysis was to uncover a causal model that

would take into account these and additional factors shaping people's responses. One specific question concerned the issue of the positive contribution made by the avatar: in what circumstances could the avatar be seen to add value?

In addition to gaining a fuller understanding of the avatar's impact, there were some additional goals in terms of the overall thesis. Firstly, to expand on some of the themes covered in the questionnaires, such as impression formation. Secondly, to refine the research questions concerning avatar fidelity: specifically, what aspects of an avatar's appearance or behaviour might contribute to the sense of being 'with' another person? Finally, to generate hypotheses to be tested in further studies.

6.3 Method of Analysis

As discussed in Section 3.2.3.3, Strauss and Corbin [SC98] describe the process of grounded theory analysis as consisting of three stages: open coding, axial coding and selective coding. They stress that although the three phases are logically distinct and serve different purposes, they often occur concurrently. This parallel use of two coding phases is most typical in the case of open and axial coding. The analysis was undertaken in three phases, as summarised in Table 6.2.

Table 6.2: Temporal overlap between the three phases of analysis

	Phase 1	**Phase 2**	**Phase 3**
Open coding	Identifying concepts in primary documents using Atlas.ti software [ATL03]		
Axial coding	Recording emerging relationships in mini-framework and conceptual network diagrams	Integration of networks	
Selective coding		Development of key categories generates hypothesis about identity of core category	Visual abstraction crystallises identity of core category

6.3.1 Phase 1: open and axial coding

The granularity of open coding can vary from word-by-word coding, to coding by sentence, and finally to coding by paragraph. The first phase began with a detailed labelling of all individual words that appeared to stand for concepts. In cases where the interview itself contained a particularly evocative word or phrase, this was used directly as an 'in vivo' code. This very

detailed word-by-word coding was used for the early stage of analysis. Later in the coding process, when concepts became more familiar, coding was often conducted by sentence or even by paragraph.

The ATLAS.ti software program was used for the open coding [ATL03]. All of the 48 interview transcripts from the two avatar conditions were included in the data pool. In ATLAS.ti, concept labels appear in the margin next to the original text. Codes can be linked with each other and with memos. This ability to cross-reference the original text with codes and memos greatly facilitates the 'constant comparative' method Strauss and Corbin stress is essential to the process of grounded theory.

It is important to avoid creating an artificial distinction between open and axial coding. As the raw data in the transcripts was labelled to identify emerging concepts, the connections between the concepts were simultaneously recorded in mini-framework and conceptual network diagrams. Though ATLAS.ti provides a tool for creating network diagrams, this feature proved cumbersome and it was simpler to record these diagrams on paper. During the coding, 493 concept labels were created (see Appendix F), and 178 diagrams were generated illustrating connections between the emerging categories. The connections in these conceptual network diagrams took one of three forms:

1. recording of *substantive* connections, for example where one concept was identified as a subcategory or property of another concept (Figure 6.1);
2. recording of the *dimensions* of a *property*: it was important to record the dimensional continua because it is at the dimensional level that connections exist between concepts (Figure 6.2);
3. recording of *causal* connections between elements (Figure 6.3).

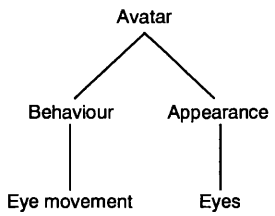


Figure 6.1: Conceptual network diagram illustrating substantive connections between categories

Depending on the complexity of of an interview passage, the mini networks could include any or all of these three different types of connections.

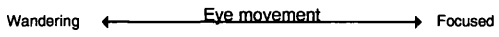


Figure 6.2: Conceptual network diagram illustrating the dimensions of a property

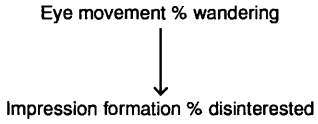


Figure 6.3: Conceptual network diagram illustrating a causal connection between two dimensionalised categories

6.3.2 Phase 2: axial and selective coding

Once all of the transcripts had been coded in this way, a second phase of axial coding began. Concept labels were merged wherever redundant. The mini-networks generated in the first phase were compared and consolidated into a large global diagram (see Figure F.1 in appendix F). This diagram soon became saturated with codes and their interconnections. However, one useful outcome was that the main 'attractors' were identified: most of the activity in the network diagram was seen to be oriented around a selection of key categories.

Next, separate network diagrams were created for each of the following key categories in order to further define their relation to subcategories and to each other:

- Context
- Experiment
- Voice
- Technical assumptions
- Communication expectations
- Mirroring
- Agency
- Association
- Avatar appearance
- Avatar behaviour
- Visual focus
- Trust
- Personal contact
- Spatial copresence
- Comfort
- Impression formation

Special attention was given to the 'avatar' categories, since there was a significant amount of complexity in terms of how the numerous properties interrelated. Memos were kept throughout to record the gradual emergence of hypotheses about what the storyline might be. This signalled the beginning of selective coding, where the analysis begins to hone in on a core category. Recall some of Strauss and Corbin's [SC98] criteria for the core category as discussed in

Section 3.2.3.3:

1. it must be central: that is, all other major categories can be related to it;
2. it must appear frequently in the data;
3. it must be able to explain variation as well as the main point made within the data.

6.3.3 Phase 3: selective coding

During the course of creating these detailed category diagrams, it became increasingly clear which categories appeared to be acting as stronger attractors. The strongest of these was 'visual focus', a category describing whether participants chose to ignore or attend to the avatar. A clear hypothesis began to emerge about its probable identity as the core category. However, the complex interconnections between categories had become so dense that it became necessary to abstract away from the verbal data to understand the overall picture that was emerging. It was also necessary to verify whether 'visual focus' actually was the category with the highest conceptual density (the most connections to the largest number of categories).

A process of abstraction was developed using colours to represent each key category, as detailed in Appendix F. This process confirmed that 'visual focus' had the highest conceptual density, and clarified its relationship to other key categories.

6.4 Analytical Findings

The previous sections presented the aims of the analysis, defined the data pool and described the method of analysis adapted from Strauss and Corbin's [SC98] guiding principles of grounded theory. This section will present the outcomes of the analysis, first by describing the key categories, and then by presenting their interrelationships in a narrative 'storyline'. It will conclude with a detailed treatment of the categories relating to avatar fidelity.

6.4.1 The key categories

The overall picture emerging from the analysis concerns the acceptance or rejection of the avatar's role in the conversation process. The process surrounding this acceptance or rejection can be broken down into three stages: In the *background* stage, people bring prior expectations and assumptions to the interaction. These shape their interpretations of the avatar's role in the *mediated experience* stage, leading people to either ignore or attend to the avatar. In the *outcome* stage, rational interpretations shape the ultimate acceptance or rejection of the avatar, as well as additional responses concerning the interaction and conversation partner. These stages are summarised in Table 6.3.

Table 6.3: Key categories

Stage	Category	Description
Background	Context	Previous experiences with virtual humans and media including VMC
	Experiment	All aspects relating to the experiment: the video-tunnel setup, task, scenario, procedure
	Communication expectations	Feedback the medium is expected to provide
	Technical assumptions	Whether it is possible to accurately convey human movement in an avatar, and whether setup is assumed to be symmetrical
	Voice	Cues available from hearing the partner's voice
Mediated experience	Avatar appearance	Static properties relating to avatar's visual appearance: photorealism, resemblance to real person, personalisation
	Avatar behaviour	Dynamic properties relating to avatar's animation: subtlety, precision, expressiveness, amount and variety of activity, degree of independence from real person's movement, degree of consistency with voice cues
	Mirroring	Degree to which participants are certain or uncertain that the avatar's animation faithfully mirrors the real movements of the person represented
	Agency	Degree of direct control the conversation partner is believed to have over their avatar's actions
	Association	Degree to which a direct association is made between the avatar and the conversation partner it is supposed to represent
Outcome	Visual focus *(core category)*	Strategy adopted by participants to either ignore or attend to the avatar, representing an acceptance or rejection of the avatar as a representation of the conversation partner
	Additional outcomes	Personal contact, spatial copresence (the sense of being in a shared space), impression formation, trust, comfort

This section has introduced the key categories. The overall process of responding to the avatars has been divided into three stages for the sake of clarity: the *background* categories partially shape the way the *mediated experience* is interpreted, resulting in perceptual *outcomes* concerning the avatar and the interaction as a whole. The following section will present the findings in greater detail using text from the primary documents to illustrate the complex connections between categories.

6.4.2 The storyline

The analysis yielded a central category, 'visual focus,' which is symptomatic of participants' acceptance or rejection of the avatar as a representation of the conversation partner. Responses to the avatar hinge on a number of interrelated factors. The avatar's appearance and behaviour are actively interpreted throughout the communication process. There is a constant interplay between the avatar's actual performance and the participant's adjusting rational interpretations about what it represents. These interpretations are shaped at the outset by a combination of communication expectations and technical assumptions.

6.4.2.1 Communication expectations

Communication expectations are shaped both by previous communication experiences and by the specific context of the conversation task. In terms of behaviour, everyday experience with face-to-face interaction gives rise to specific requirements in terms of visual feedback. In cases where people have also experienced videoconferencing systems, the expectation is that a visual image should provide them with sufficiently precise dynamic feedback to give them an accurate 'reading' of the other person. The avatar is therefore expected to provide cues about the partner's reactions and emotional state. These cues enable participants to adjust their negotiation strategy and get a measure of how the conversation is progressing. The limited behavioural capacity of the avatar fails to meet these needs: *"You know, when you're face-to-face then an awful lot of it goes in the face, doesn't it, and in the hands and those kinds of nonverbal communications. Really I would say that I was getting very little nonverbal communication from the virtual human" (C2G4A).*

The avatar is seen as providing insufficient feedback about the partner's reactions. What little feedback there is seems to offer only vague suggestions rather than a clear message: *"You get sort of an impression of the facial expression from the virtual presence" (C2G5B).* Certainly it is not sufficient to get a reading of the partner's emotional state: *"I would have said it was, you know, totally unlike a face-to-face conversation because you don't get the range of expressions and you can't see emotions or anything like that" (C2G3B).* Greater quantity and clarity of feedback are therefore needed.

Communication expectations are also shaped by the experimental brief, which describes a tense emotional scenario that calls for contained and focused interaction: *"There seemed to be too much injected random motion. I would have expected the natural slight movement but he was moving around a lot, which in a normal situation you would have said, 'don't you take this seriously?' Because he didn't seem to be that serious. We only had ten minutes, we would be really, you know, really locked into each other. The movement would be extremely small. You*

might get where someone is thinking, they might look away and come back, but nothing more than that" (C3G4A). The context is therefore crucial in defining expectations. This scenario calls for the kind of staid behaviours also appropriate in a business meeting. A lighter social context would, in contrast, call for more lively and expressive behaviours: *"I suppose maybe you'd look at it differently if you were using it for business and if you were using it for social. I think if I was talking to a friend I'd want more hand gestures but maybe if it was for business use, then you wouldn't expect someone to be waving at you the whole time" (C2G8B).* The amount of liveliness appropriate is therefore partly dictated by the communication context.

6.4.2.2 Technical assumptions

Technical assumptions, like communication expectations, impact directly on how the avatar's behaviours are interpreted. These assumptions are informed partly by a general experience with mediated communication and by an awareness of the techniques used for driving avatar behaviours. Appreciation of the technical difficulties involved in tracking full body and facial behaviour leads to scepticism about whether the avatar can, even in principle, accurately represent a person in real time: *"I know you can do eyetracking. I don't know whether you can do... I sort of smiled a bit you know, so it's not going to be easy to pick up, let's put it that way" (C2G7A).* This scepticism is partly fuelled by questions about how one's own behaviour might be captured and portrayed to the other person.

This doubt regarding the likelihood of accurate tracking, coupled with the lack of visible tracking equipment in the setup, leads to a sense of uncertainty about whether one can be seen by the other person: *"I was a little unsure at first whether that was mimicking any of my movements, were there any cameras around here that could tell when I bowed my head and whether he was seeing the top of the character's forehead and things like that, so I was a bit conscious of that" (C2G12B).* The assumption, whether conscious or not, is often that one's own behaviour is not being portrayed to the conversation partner at all: *"I assumed that she couldn't pick up on my facial movements but actually maybe that wasn't the case. I didn't think about that. I made that assumption" (C2G8B).* This in turn leads to the belief that one is being neither directly nor indirectly observed by the conversation partner, meaning that it is not necessary to visually communicate interest and attention by looking at the screen. Whereas videoconferencing triggers social norms of politeness, here the reaction is different: *"I was sort of leaning around looking around here which I wouldn't do if I was talking to the person in the room or even in a videoconference because that in itself would give the wrong impression" (C2G9B).* There is a widespread conviction that the avatar is not owed the same kind of polite attention as a real human.

There is also a widely-held assumption that the majority of technical setups are symmetrical: *"I actually thought he's probably got the same setup as me and so I didn't think, oh I wonder, but I just automatically thought: right, if he's having this kind of conversation with me he's probably got the exact same equipment" (C3G1A).* Symmetry functions as a catalyst in undermining the relevance of the avatar in conversation, and in many cases leads people to stop paying attention to it altogether: *"I think I sort of didn't give that much respect to it. Knowing that he couldn't see me, I didn't give that much respect" (C2G12B).* Many rationalise that 'respecting' the avatar as they would a video image does not communicate attention and respect to the other person: *"You know it's an avatar, you know that looking at the screen doesn't help her see you" (C2G11A).* Technical assumptions therefore have a direct bearing on whether the avatar can be perceived as a mediator between the two participants. This idea of 'knowing it's an avatar' and not a real person comes up repeatedly and underlines the importance of people's rational interpretations in accepting or rejecting the avatar.

Given that the avatar is computer-generated, there is a great deal of uncertainty as to what it represents. Unlike a video image, it needs to earn the user's belief and trust that it is a visual representation of the other person. One of the key ways in which the avatar can achieve this is by accurately reflecting the person's real-life actions.

6.4.2.3 Mirroring

'Mirroring' is a key element in people's responses to the avatar. It is a rationalisation about the degree to which the avatar is portraying the partner's actual movements. Given that those real movements cannot actually be seen, assessments are based on whether the behaviour appears appropriate in the given context, and whether it reflects the information transmitted by the partner's voice. Mirroring can range from certain to uncertain, and from faithful to unfaithful (Figure 6.4). Faithfulness refers to the accuracy with which the avatar represents the user's real life actions. Certainty refers to knowledge about whether the mirroring is faithful or unfaithful. In the absence of explicit information, participants make informed guesses about mirroring.

Technical expectations impact directly on people's certainty about mirroring: *"It may or may not represent the real behaviour of your conversation partner at the other end. I think it could be a distorted image we have been watching. There could be some mismatch between the image and the real thing. I don't know, maybe you have some fantastic techniques to duplicate whatever your partner's mood is, but I doubt it" (C3G2B).* In the absence of explicit information about how the avatar works, rationalisations about the avatar's mirroring are shaped both by its behaviours and by assumptions about what is technically possible.

Given that the two participants are not in the same physical room, they can only speculate

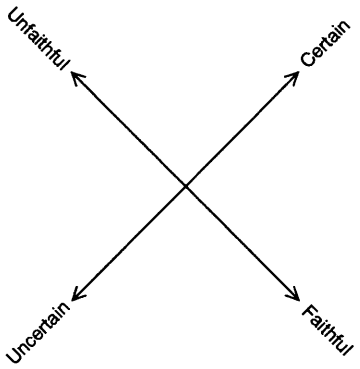


Figure 6.4: Dimensions of 'mirroring'

about their conversation partner's real actions. As mentioned above, people expect more sombre behaviours in this charged interaction context than they might in a purely social setting. For example, when the avatar's eye movements appear too flitting they fail to reflect the focused attention received through the audio exchange. If the voice appears to express attention and the avatar continually looks up in a disinterested fashion, the mirroring is assumed to be inaccurate: *"From the phone conversation, if that person was really in front of me, I would have imagined they were looking at me because the audio conversation sounded very good and I could not imagine that sort of interaction if they weren't looking at me" (C3G2A).* In this case, there is a clash between the avatar's actions and the actions inferred by listening to the voice.

The discrepancy between voice cues and avatar behaviours heavily impacts on interpretations about mirroring and ultimately undermines the avatar's credibility. The result of unfaithful mirroring is that the association between partner and avatar is automatically compromised. Poor lip synch provides concrete evidence of a clash between visual and aural cues: *"As soon as I saw that it was slightly adrift I basically switched it out. I tend to do that with badly dubbed films and things like this, or where the sound synch goes off. You know, I mentally switch it out and just tune in to audio. I think that's probably what I did here quite a lot" (C2G6A).* Poor lip synch therefore provides clear evidence of inaccurate mirroring. However, it is sufficient for mirroring to be uncertain in order for the association to be undermined. Doubts about the faithfulness of animation are sufficient to call into question the avatar's role.

6.4.2.4 Agency

An alternative to mirroring a person's real actions is to represent the meaning they intend to put across. While mirroring refers to the reflection of a person's real physical actions, agency refers to the representation of their intentions. The agency of a virtual human refers to the intelligence that is driving its actions. This intelligence can either be human (as in the case of

user-driven avatars), or artificial (as in the case of embodied agents). Human agency presumes a direct and deliberate connection between the partner's intentions and the resulting behaviours of the avatar. This necessarily strengthens the association between the avatar and the person represented. However, the ability to control an avatar undermines trust in the other person because it admits the possibility of masking one's real emotion by 'lying' through the avatar: *"At the end of the day I suppose you can contrive what the figure does. I'm not sure I could trust someone talking to this, not if I haven't met them before in person and get an idea of who they are, their motivations. That's true in any situation, I suppose" (C2G6B).*

Agency and mirroring are therefore distinct approaches to representing a real person, and consequently result in different reactions in terms of both association and trust. Faithful mirroring results in a high level of association as well as trust in the avatar's behaviours because it gives an accurate reading of what the person is actually doing. Agency, on the other hand, results in a high degree of association between person and avatar, but a lower level of trust because the avatar can be controlled to 'mask' real emotions.

6.4.2.5 Association

Both mirroring and agency are principally tied to the avatar's movement. If jeopardised, both can undermine the association made between partner and avatar; in other words, if the avatar's behaviours are not seen as representing any aspect of the partner's intention or real activity, they will cease to have any communicative meaning and will be ignored. However, association is also a function of other elements of the experience, including the partner's voice, the scenario, the avatar's appearance, and the fact that the participants saw each other briefly prior to the start of the experiment.

In terms of the avatar's appearance, the scenario describes a meeting between a mayor and baker, both parents of adult children. This leads to an assumption that the conversation partner will be of a certain age, social stature and appearance. The fact that both the male and female avatar had the appearance of a twenty-something contradicts the expected age of the partner within the context of the roleplay: *"She was younger than I thought she was going to be. I was expecting her to be mid-fifties, being a woman a mayor and having a son, and she looked more late teens early twenties, which distracted a little" (C3G3B).* This is particularly problematic for participants age 40 or over, who experience an additional discrepancy between the apparent age of the avatar and the age suggested by their partner's voice: *"Somehow I felt that the voice didn't go with what I see there. You know, the age seems wrong and perhaps the build or something like that. That doesn't really seem to match the kind of voice and the expectation I've got from reading the brief" (C2G4A).* Expectations about age and physical appearance also

clash with other details, including the avatar's youthful and casual apparel: *"It's difficult with the attire. I don't really know, you know, this is the mayor I'm speaking to of wherever I am, and it's unlikely to have a t-shirt on" (C2G10B).*

These considerations concern the tension between the avatar's appearance and the appearance expected in the context of the roleplay. The brief real-life meeting prior to the experiment only serves to exacerbate this tension because participants are aware of the real appearance of their conversation partner. Seeing the other person, even if only for a few moments, appears to play a major role in undermining the association between avatar and conversation partner: *"I didn't really consider him a virtual version of the other party because I mean I've already seen the guy, so you know it just seemed as though there was something on the screen making the lip movements that coincided with his" (C2G3B).*

A greater resemblance between the avatar and the person represented would also affect other aspects of the experience, including the motivation to look at the screen: *"If the avatar was her, like it was a physical representation of her with blond hair and whatever, then I think I would [look at it] because obviously I would know that she could see me as well and she could see what I was doing" (C2G11A).* In other words, the fact that the avatar does not reflect the person's real-life appearance also affects assumptions about whether or not one can be seen by the conversation partner. This assumption seems not altogether rational, because objectively the appearance of one's avatar has nothing to do with one's ability to see the other person. However, it is important to recall that the logic of the user does not necessarily reflect the way the technology actually works. In this case, resemblance is tied to association, and greater association means that the avatar needs to be treated with greater respect.

It is significant that association in terms of appearance translates into an assumption about mirroring. The logic seems to be that if the person's appearance is accurately represented, then their behaviours must be also. Coupled with an assumption about the symmetry of the setup, this means that one's own appearance and behaviours are also accurately represented, therefore the same social norms of politeness that operate in VMC and face-to-face communication should come into play.

This seemingly illogical link between the appearance of the avatar and the sense of being observed illustrates the complexity of the interrelationships between assumptions. They seem to have a domino effect whereby the failure to satisfy one condition leads to the belief that other logically separate conditions also cannot be satisfied. It can take only one factor, in this case appearance, to undermine the relevance of the avatar and lead to its rejection. The common reaction is to ignore the avatar if it isn't seen as representing the other person: *"I probably took*

less notice of it because I knew it didn't look like her" (C2G11B). The strategy then is to rely on the input from the audio connection, since that is believed to be a direct link to the real person: *"I was thinking, well look, that doesn't look anything like him, so I'm going to ignore it and just listen to what he's saying" (C2G4B).*

An additional factor in the technical setup that further detracts from the association between partner and avatar is the physical displacement between the avatar seen on the screen and the sound heard in the headphones. This often results in a logical dissociation between the two: *"I think because the voice was through the headphones and not coming from behind as you normally get with a TV it felt a little bit disjointed, not connected . It was a representation, not the person speaking" (C3G3B).* This audiovisual disjoint deepens the dissociation between the avatar and the voice of the conversation partner.

The breakdown in association can occur at different stages of the experience, and for any combination of the above factors. In some instances the avatar is rejected at the outset because of its lack of resemblance to the real person. In others the association remains ambiguous, buoyed by fortuitous moments when the avatar's behaviour aligns with cues from the voice: *"Smiling and the laugh at one point... She looked as though she was laughing. I thought that was probably reflecting her behaviour" (C3G3A).* This is an example of a fortuitous moment when the avatar's behaviour is in line with cues from the voice. However, the fact that this alignment is only occasional introduces an element of uncertainty, causing participants to monitor the avatar closely for signs of mirroring.

Once the association between avatar and partner is undermined, it is very difficult to regain the participant's belief in the avatar's relevance or usefulness in the conversation. It is as though the damage is irretrievable: *"The connection had gone. Basically I lost the connection, then I was just talking to the microphone" (C3G1A).* Once the connection is broken, then visual focus is withdrawn from the avatar: *"Partly I guess because I wasn't always focusing on the avatar. You are aware that they are two sort of separate entities" (C2G2B).*

The decision to either ignore or attend to the avatar is emblematic of its acceptance or rejection as a viable representation of the conversation partner. It must be stressed, however, that several positive communication outcomes are possible independently of the avatar's success or failure in playing a role in conversation. These include impression formation, personal contact, co-location, and trust. Although the avatar does have some impact, these positive outcomes are mainly due to the audio interaction. Experience with telephone conversations accustoms people to gathering information from the voice in the form of verbal content, pauses and voice tone, making participants adept at gathering feedback from the verbal interaction.

6.4.2.6 Personal contact

The sense of 'being listened to' by a 'real person' (rather than a computer) contributes significantly to the sense of personal contact. The time pressure on the task limited participants' ability to break the ice through banter, making the conversation less personal than it could have been. Banter aside, it was stressed that the sense of personal contact was no different than what could be experienced on the telephone, with the avatar making little if any positive contribution: *"I think the personal contact came through the conversation rather than the avatar" (C2G12A).* Rather than contributing, the avatar detracted from personal contact in two key ways. Firstly, because of its generic visual backdrop; rather than appearing in a specific visual context such as the mayor's office described in the brief, it was backed by a flat, coloured backdrop that was not particularly evocative and made the interaction less personal. Secondly, through its generic appearance: *"I think it would have been more important if it did actually look something like the person I was speaking to. As it is, I imagine you probably see the same avatar whoever you're speaking to and that would actually make it seem more impersonal" (C2G2B).*

On the other hand, in some cases the avatar did seem to make a positive contribution: *"I definitely felt like I had more contact than if I was just on the telephone to him, but it's more a 'something rather than nothing' situation, more than you know this is a believable character" (C2G10B).* In spite of a lack of resemblance to the real person, the avatar nonetheless served as a placeholder for the conversation partner. Many stressed that having the avatar there served as a visual aid to help them focus on the conversation and on the person they were speaking to: *"It represented the person but because it wasn't the actual image of the person I was talking to, it's more of a visual stimulus to encourage concentration, and it's helpful to have something to look at when you're looking for responses as well" (C3G3A).* Given the lack of feedback from the avatar, it does little to enhance the sense of personal contact. However, its mere presence was helpful for some participants in giving them a keener sense of interacting with their partner: *"I think it certainly added something, so it's midway between having the person there and having nothing at all" (C2G1A).*

6.4.2.7 Spatial copresence

While the sense of personal contact is affected both by the audio and by the avatar, the sense of spatial copresence (being in the same space) is primarily a function of the setup. The distance from the monitor seems excessive and serves to underline the physical distance between the partners, reminding them of the fact that they are in separate physical spaces: *"I guess in a face-to-face situation you'd probably be a bit closer than that, I don't know how far away that is. Notionally he looks like he's sort of beyond the other side of the room almost" (C2G7A).*

In some cases the sense of spatial copresence is boosted by the audio: *"You know though that's kind of contradictory it did feel like I was in the same room. But I did have hardly any visual image at all, I just went solely on the voice" (C2G3A).* The contribution of the voice is largely due to the high quality of the audio connection: *"I know he only is in the next room but he sounded very very close because of the headphones" (C2G3A).* Overall, however, partly because of its failure to satisfactorily represent the other person the avatar itself does not make a contribution to the sense of shared space: *"It didn't make me feel there was a presence of the other person there" (C2G1B).*

6.4.2.8 Impression formation

In the case of perceptions of the conversation partner, impressions were again primarily formed through the audio. The voice gave participants a reading of their partner's sincerity, diplomacy, cooperativeness and interest in the conversation. While several positive impressions were formed through the voice, the avatar was mainly responsible for negative impressions. This is not surprising given that participants had a head-and-shoulders view of an avatar whose movement was restricted to eyes, head and mouth. Very little could therefore come across in the way of emotional expression. However, the primary responsibility for the negative impressions lies with the eye animations. These were a product of both direction and frequency information.

In terms of direction, when the avatar looked upwards it communicated a sense of disinterest in the conversation. As mentioned above, when the boredom communicated by the avatar contradicted the interest communicated by the partner's voice, the rational response was to ignore the avatar's behaviours as inaccurate. However, despite the avatar's obvious shortcomings and resulting rational interpretations about its inability to mirror real movement, people's responses were not always entirely rational. On some level, the eye animations affected the way the partner was perceived: *"It's funny, the only thing is I'm imagining the other person to have eyes that go all over the place, that maybe this other person doesn't like to look at the person he's speaking to. Yes, I mean there are some people sometimes who are shy to look straight in the eyes of the person and then they look up there and they look up there. I got that impression sometimes because the eyes were up in the air" (C3G2A).* In this case, the fact that the eyes looked away led the participant to assume the partner was shy. It is significant that although the avatar's behaviours do not appear reliable, they can nonetheless have a potentially negative impact on impression formation.

In terms of the frequency of movement, they eyes were seen as 'wandering.' Partly this reinforced the sense of disinterest: *"They're just showing a look of disinterest. In a way I think that was the most obvious thing, the fact that it's looking around as if he's not really looking*

at you" (C2G10B). The fact that the avatar was looking around also gave the impression that it was not giving sufficient weight to the serious situation at hand: *"Maybe that's because of the situation. I don't know, but I felt it was not focused enough because I think in this kind of conversation you want more kind of eye contact so if that was the other person I would rather have it looking directly at me" (C3G1A).* Again, the context shapes views about the appropriateness of given behaviours. This wandering eye movement is viewed as particularly inappropriate when the person is speaking, because it gives the avatar a 'shifty' air.

6.4.2.9 Trust

Earlier it was mentioned that mirroring can range from certain to uncertain, and from faithful to unfaithful. Uncertainty was sufficient to undermine the relevance of the avatar, but also to raise questions about the conversation partner's sincerity: *"Whether that means that the person at the other end is actually doing that or whether they're sort of being evasive or whatever I don't know" (C2G7A).* The avatar's behaviour does not directly result in a negative impression, however it raises uncertainty and does not enhance the connection between conversation partners. However, the rational interpretations usually take the upper hand. The end result is that the avatar's eye movements do not make the conversation partner appear untrustworthy; rather, they made the avatar appear untrustworthy as a faithful representation. The trust for the other person is fostered, again, thanks to the voice: *"The voice gave the trust, not the picture" (C2G12B).* The positive impressions about the partner's diplomacy and sincerity play a significant role. The avatar does not significantly affect the perceived trustworthiness of the partner simply because it does not give sufficient visual information to form any solid impressions of the other person: *"I think it's got quite a way to go yet in terms of really seeing what a person's thinking, telling whether a person's lying or not. I think it's got some way to go. Mostly you're relying on the voice information" (C2G4A).*

Several ways have been mentioned in which trust in the avatar can be undermined. The only way in which the avatar could hypothetically be trusted would be if it accurately mirrored the other person's actions: *"If you're not looking at the actual person then how could you actually have that guarantee that what you are looking at was doing what they were doing? Because you know it might be just pretending that it was" (C2G7A).* The notion of a 'guarantee' flags an issue of central importance throughout the analysis: the continual dialogue between rational beliefs and the avatar's performance. In the hypothetical instance of a highly photorealistic avatar capable of accurately mirroring a person's behaviour in real time, it would still need to offer some assurance of its integrity as a representation of the real person: *"Only after people have had some knowledge that this avatar can really represent the opposite party [would it be*

useful]. For this thing to be of some help I think you had better do some education, that is convince people that thing can really mimic what you are talking or mimic your mood or behaviour. Otherwise, for example I am a technical person, now I know that these sorts of things are still in a very primitive stage. Because of this I immediately started paying very little attention as I said, cause I know there is a huge discrepancy between what you are seeing and what the other real person was doing" (C3G2B). In other words, it is not sufficient for the avatar to resemble the person and mirror their actions. People would also need to be persuaded that this is the case; they would need a guarantee.

When confronted with ambiguity or uncertainty, participants profess an inclination to believe whichever aspect of the communication is more human. If presented with live video of a human face coupled with a computer-generated voice, they are more likely to believe the video. Here, when presented with a computer-generated person coupled with an obviously human voice, they place their full reliance on the audio.

6.4.2.10 Core category: visual focus

Rejection of the avatar occurs due to interconnected factors that, working together, undermine its credibility as a representation of the other person. Rational interpretations are made of the avatar's performance. These are based partly on technical assumptions and communication expectations that people have prior to the interaction. Once the association between the avatar and the person represented is jeopardised, trust in the avatar as a meaningful representation is also lost and the avatar ceases to have relevance in the conversation. The common strategy is then to stop paying visual attention to the avatar and to focus instead on the audio interaction.

This storyline has focused on the interconnection of factors contributing to the acceptance or rejection of the avatar as a meaningful factor in the communication process. The following section will focus on the properties relating to avatar fidelity, with a view to uncovering ways in which performance can be improved for the purposes of mediated communication.

6.4.3 Avatar fidelity: appearance and behaviour

Two principal avatar categories feature in this analysis: appearance and behaviour. Appearance concerns the avatar's static properties, namely the geometry and textures that give it a visual identity. Behaviour concerns its dynamic properties, namely its animation. Both appearance and behaviour are actively interpreted, and both have an impact on the perception of the avatar. This section will focus on the avatar categories in order to illustrate how they affect some key responses: mirroring, trust and association.

6.4.3.1 Visual fidelity

There are three properties in this analysis relating to the avatar's appearance: realism, resemblance and personalisation (Figure 6.5).

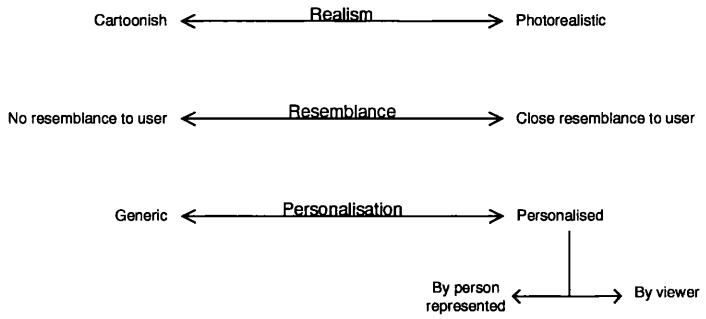


Figure 6.5: Properties relating to the avatar's appearance

The appearance of an avatar is often described as ranging in realism from cartoonish to photorealistic. The avatar was generally thought to be relatively cartoonish, and this in many cases undermined the association. It was seen as a puppet rather than as a representation to be taken seriously. Some participants argued that had it been more realistic, they might have associated it more closely with the person they were speaking to.

Additional properties have been uncovered in this analysis as being directly relevant to the acceptance or rejection of the avatar: resemblance and personalisation. The avatar in this experiment was generic, although participants were not aware that their avatar was visually identical to their partner's. However, they were able to see that the avatar in front of them did not resemble the real person they had seen on their way in. An avatar's appearance can range from providing an accurate likeness of the real person to bearing no resemblance whatsoever. Here, lack of resemblance played a significant role in the dissociation between partner and avatar, and sometimes resulted in the avatar being rejected at the outset.

Aside from the lack of resemblance, the assumption that the same generic avatar was used to represent a number of individuals also undermined the sense of personal contact. An alternative approach to aiming for accurate resemblance is therefore to personalise the avatar so that it is at least tailored to each individual. Personalisation can be approached in at least two ways. The avatar can be personalised by the person represented, or alternatively by the viewer. In either case, the ability to choose from a selection of avatars or, better still, to customise an avatar's features and clothing would both be valid steps towards personalisation. Whether the personalisation is in the control of the person represented or the viewer will likely have different implications, and the resulting impact on trust and association, as well as the general role of the

avatar, opens avenues for future research.

6.4.3.2 Behavioural fidelity

The behaviour category is subdivided into two subcategories: quantity and quality.

Quality of behaviour: There are three principal properties relating to quality of avatar behaviour: subtlety, precision and expressiveness (Figure 6.6).

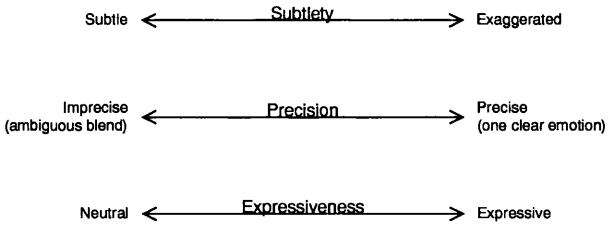


Figure 6.6: Properties relating to quality of behaviour

The avatar was overwhelmingly seen to be too neutral in expression, and its failure to communicate any real emotional information was seen as one of its major downfalls. Sometimes it appeared to give slight indications of expression, but these were so muted that they failed to transmit a clear meaning. *Subtlety* refers to how exaggerated an avatar's expression needs to be in order to get its meaning across, and in this case the expressions were judged to be insufficiently intense to be noticeable. This is to be expected, because in reality the avatar was not putting across explicit facial expressions. Instead, expressions were inferred by participants from a combination of eye, head and mouth movement.

Precision refers to the ability to put across a single clear emotion, as opposed to a blend of emotions. The avatar is already seen as an ambiguous representation of the conversation partner. Therefore precise, unequivocal emotional expression is seen as a priority if the avatar is to make a meaningful contribution to communication.

Quantity of behaviour: The properties relating to quantity are closely tied with activity over time. These are variation, independence, consistency with the voice, and amount of activity (Figure 6.7).

Variation concerns the degree to which the avatar's movement is repetitive. Loops in the animation made it appear robotic and unlike a human. The *independence* of the avatar refers to the degree to which its movement is autonomous from the person's. In this experiment, the fact that it could be seen to move on the screen before the beginning of the conversation make it appear to be a separate entity. Any behaviours that are inconsistent with the voice compound this impression. As illustrated in the previous section, the amount of *activity* also has a significant impact on how the avatar is perceived. In this case, its eye movements and *"injected*

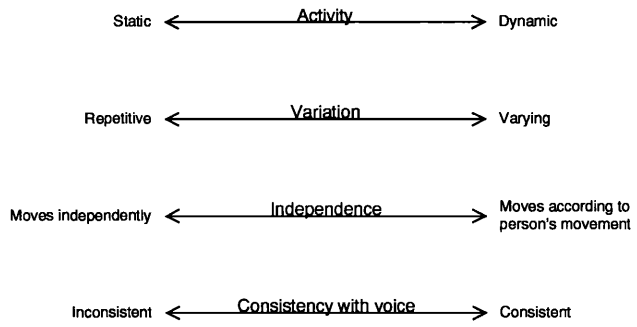


Figure 6.7: Properties relating to quantity of behaviour

random motion" (C3G4A) make it appear too jumpy and shifty. This is not desirable, particularly in a negotiation context where trust between participants is paramount. A value judgement is made about whether the amount of activity is sufficient given the communication context. In this particular instance, given that the situation was serious, the amount of activity was seen as excessive, and communicated a lack of concern. Value judgements are also made about the general appropriateness of behaviours given the communication context. In this particular case, it would not be appropriate to smile too much, given the gravity of the situation being discussed. It would also be inappropriate to look around too much, as this communicates a lack of both trustworthiness and interest. Therefore context shapes not only the quantity, but also the quality of expressions deemed appropriate.

One property that belongs to both quality and quantity concerns balance across features. When asked about the behaviours they noticed, participants mentioned the eyes and the mouth. The fact that some features were more animated than others created an unnatural effect, because real human behaviour is a gestalt of movement and is not usually broken down into individual features. This experiment was designed to focus on a single behaviour, eye gaze. However, for an avatar to be convincing there should be an equilibrium of animation across all facial features. There is an interaction effect between quantity and quality of animation (Figure 6.8).

In this particular communication context, insufficient precision coupled with an excess of activity makes the avatar's behaviour appear unclear at best, and nervous and aggressive at worst. A smaller amount of precise behaviours would contribute more to communication by putting across unequivocal meaning and by suggesting a relaxed, natural pose. This serves as a caution against giving an avatar 'liveliness' without studying the exact impact of these added behaviours, since even small shifts can result in a negative impression.

To summarise, some guidelines emerge from this analysis on the combination of properties that would increase the avatar's contribution to communication. The avatar's appearance should

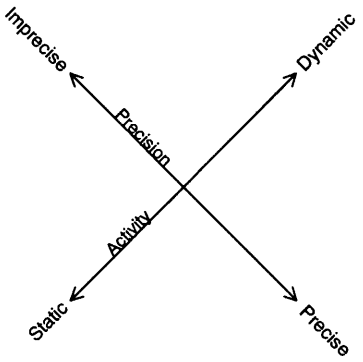


Figure 6.8: Intersection between quantity (in terms of activity) and quality (in terms of precision) of animation

be personalised, and ideally should resemble the real person in order to strengthen association. The appropriate amount of expression is partially dictated by the communication context: business meetings and serious discussions call for more contained behaviours, whereas social contexts allow for more dynamic and lively behaviours. The avatar's movement should not be visibly repetitive, and it should not occur independently of the activity of the user. Ideally it should be consistent with voice cues. Also, in terms of the eye animations it is preferable to avoid too many upward movements of the eyes, since these express disinterest and detract from the conversation, while undermining the believability of the avatar. The avatar's expressions should be sufficient to convey emotional information. Also, these expressions should be precise and discrete in order to put across a single, clear meaning rather than an ambiguous blend of emotions. However, the degree of subtlety of the expressions needs further investigation; it is not clear how intense or exaggerated an avatar's expression must be in order to come across clearly. Finally, there should be a balance in the amount of activity across different features. Excessive movement in the eyes, coupled with insufficient movement in other features, creates an unnatural effect.

6.4.3.3 Predictions

As discussed in Chapter 3, grounded theory was originally developed with a view to hypothesis generation. The goal of analysis is to move beyond anecdotal description to the discovery of causal models with predictive power. In principle, it should therefore be possible to draw some hypotheses from the analysis presented. This chapter concludes with a list of testable hypotheses for future research.

Concerning visual fidelity:

1. Avatars with a known *resemblance* to a real person should enhance the sense of *personal*

contact;

2. Avatars with a known *resemblance* to a real person should result in a greater degree of *association* between the avatar and the person represented;

3. Similarly, *personalised* avatars should result in a greater degree of *association*.

Concerning behavioural fidelity:

1. Knowing that the conversation partner has a great degree of direct control over the avatar's actions *(agency)* should result in increased *association*, but a lesser degree of *trust*;

2. Knowing that the avatar's actions *faithfully mirror* the real movement of the person represented should result in increased *association* as well as *trust* in the avatar;

3. Avatars with a smaller number of clear expressions are more likely to result in positive perceptions of the interaction than those with a greater number of unfocused expressions;

4. Mirroring is more likely to result in higher trust than is agency.

Another hypothesis to emerge from this analysis is that personal contact does not necessarily correlate with spatial copresence (the sense of being in a shared space). Although both of these constructs logically relate to copresence, they may not share the same root causes. This signals a need to further define the indicators of copresence. An additional hypothesis is that personal background, in terms of previous exposure to different media, will shape expectations of mediated experience, and therefore perceptions of the avatar's role.

Not all the emerging hypotheses concerning visual and behavioural realism were tested in the subsequent experiments, primarily because several of them concern the priming of participants. The aim of this research was to investigate participants' spontaneous perceptions of different aspects of avatar fidelity in the absence of explicit knowledge about their functionality. Nevertheless, both personal contact and spatial copresence were investigated in the two subsequent experiments, as was the impact of prior exposure to related technologies.

6.5 Chapter Summary

This chapter has focused on the analysis of the 48 interviews with participants in the two avatar conditions: *random-gaze* and *inferred-gaze*. It began by defining the goals of the analysis and describing the general themes identified by reading the interview transcripts. It then described how Strauss and Corbin's grounded theory method [SC98] was adapted by using colour to abstract from the categories and verify their interrelationships as well as the central position

of the core category. Finally, it presented the findings and discussed how the balance between various aspects of visual and behavioural fidelity might be adjusted to enhance the avatar's potential role in communication.

The analysis reflects the responses of participants who had an isolated experience of avatar-mediated communication. Several limitations define the scope of this analysis. Their conversation partner was represented by a 'generic' avatar that did not resemble the real person. Participants had a non-immersive interaction, experiencing a head-and-shoulders view of the avatar on a video-tunnel monitor in front of them. Furthermore, the avatar's behaviours were limited to eye, head and lip movement, with no explicit facial expressions. The avatar's ability to provide feedback was therefore severely limited. In addition to this, the role-playing negotiation task meant that high demands were placed on the avatar's ability to provide emotional feedback. Overall this was an unforgiving context in which to test its performance. However, this was deliberate since the aim was to explore people's expectations and responses without making allowances for the technical limitations of the system. One of the goals of the analysis was to identify priorities for improving communicative avatars. For this reason participants were not briefed on how the avatar worked, or which aspects of their partner it represented.

Nowak and Biocca ask the question: "When bodily cues are computer-generated and not natural, will people rely on them for person perception and social judgment or ignore them because they are completely fabricated and, therefore, untrustworthy?" [NB01, p.12]. The grounded theory analysis suggests that people will attend to whichever cue they believe is the most trustworthy representation of the other person. The degree to which they perceive the avatar's behaviours as enhancing the conversation hinges on a number of interrelated factors. These include external factors relating to avatar fidelity and the interaction context, as well as internal factors such as expectations formed by prior technical experience.

Overall the analysis points to the importance of managing people's expectations by explaining how the technology works. In the absence of any explanation, people's rational responses are shaped by their expectations of visual communication as well as by their assumptions about what is technically possible. Often these assumptions will work to the detriment of the avatar by undermining the role it plays in communication. Having said this, it is also true that people's responses are not exclusively rational. In some cases, the avatar's behaviours will lead people to form impressions in spite of the fact that they rationally believe it does not reflect what their partner is actually doing. In this eventuality, it is preferable for the avatar's actions to avoid misrepresenting the person's attitude. Specifically in the case of eye gaze, this analysis suggests that an overabundance of movement coupled with too many upward glances results

in an impression of disinterest and evasiveness. If this impression contradicts the impressions formed from the voice, the association between the avatar and the person it represents will be further undermined. Once the association is broken, it is very difficult if not impossible to re-build. The way in which an avatar is likely to be accepted and to maximise both association and trust is by guaranteeing that it can mirror a person's actions in real time. However, this poses significant technical challenges. In the absence of faithful mirroring, it would be beneficial to inform users as to which exact aspects of a person's real actions or intentions the avatar is capable of reflecting in real time.

Chapter 7

Discussion

The overarching goal of the three experiments presented in this thesis was to investigate the impact of avatar fidelity on people's experiences of social interaction in CVEs. Chapters 4, 5 and 6 presented research findings from three experiments investigating the perceptual and social effects of three key aspects of fidelity: behavioural fidelity in terms of both *responsiveness* and *eye gaze* behaviour, and visual fidelity in terms of *photorealism*. This chapter will summarise and discuss the findings. For convenience, Appendix A summarises and compares the goals and design of the three experiments.

7.1 Experiment on Responsiveness

As discussed in Section 2.4.4, responsiveness is believed to be a key component of avatar fidelity [SB02]. Heeter proposes that responsiveness in both objects and virtual humans should directly enhance the overall sense of presence by making the user feel like an acknowledged part of the VE [Hee92]. The experiment presented in Chapter 4 focused on the impact of basic responsiveness on a range of social responses to humanoid agents. Participants explored a virtual library containing five seated 'readers.' The agents were visually identical in all four conditions, but their behaviours were modified across four conditions to range from *static*, to *moving* to *responsive*, and finally to *talking*.

The animation approach taken was comparable to that of Pertaub et al. in their experiments on fear of public speaking [PSB01b, PSB01a], where the agents were given a selection of behaviours to generally indicate whether the audience was responding negatively or positively to the participant's talk. While the experiments on eye gaze aimed at isolating a single behaviour, this experiment on responsiveness sought to investigate the impact of a more diffuse collection of behaviours suggesting that the agents were either studying undisturbed, or aware of the participant when approached in the *responsive* and *talking* conditions. Responsive behaviours included sitting up straight, turning to face the participant, and engaging in mutual gaze.

The aim of the experiment was to investigate whether basic increases in responsiveness could contribute to participants' sense of interacting with people. The primary responses of interest included their sense of being in the company of people (copresence), the level of awareness attributed to the agents, and their reported behaviour in response to them. Additional responses included the sense of presence, and objective psychophysiological responses in the form of heart rate and EDA.

7.1.1 Copresence and presence

The lack of consensus on the definition and measurement of both presence and copresence was discussed in detail in Section 2.5. The definition and measurement of copresence developed over the course of the three experiments in light of ongoing analysis as well as related publications by other researchers. In this experiment, the measure of copresence consisted of items relating to the sense of personal contact and the sense of interacting with other people as opposed to a computer interface.

When considered on its own, copresence is significantly lower in the *static* condition than in the three conditions where the agents moved. Several subjects in the *static* condition reported that they had a low sense of personal contact because the agents did not move, and therefore it did not occur to them to attempt any form of interaction: "Since there was no movement that was discernible, they were people but they were simply objects there, so any type of interaction was impossible. So it was always in a sense like a still painting. If you take a look at that you will see the still painting or a picture and it is not a real interaction." The sense of personal contact was also hindered in the *moving* condition by what several participants described as an almost voyeuristic experience. The fact that the agents moved but did not respond made some participants feel 'invisible' and 'ghostlike,' and unable to engage in two-way interaction of any form. Thus the lack of responsiveness in the agents coloured the experience in the virtual environment as a whole by making them feel they could not affect their surroundings. This echoes the suggestion made by Heeter that unresponsive characters will lead to a sense of disconnectedness from the VE as a whole, therefore directly affecting presence [Hee92].

However, the results do not fully support this notion, since the agent conditions did not appear to directly affect reported presence. However, when considered in relation to participants' level of computer experience, the picture changes considerably, with the *talking* condition resulting in a significantly higher presence response. There is an inverse relationship between presence and computer usage, with more experienced users being less likely to feel present in the VE.

The same pattern emerges for copresence. When considered in relation to computer usage,

then the *responsive* condition results in a significantly higher copresence response. In addition, a greater degree of computer experience is associated with a decreased sense of copresence. This seems to imply that experienced computer users are less likely to suspend their disbelief and respond to the agents as people. Several participants who reported playing games tended to refer to the agents in a more detached way in the interviews. They reported trying to provoke a 'reaction' from them, rather than treating them as people. One participant in the *responsive* condition explained: *"Normally meeting new people you would keep more of a distance... I was closer to them than I would be with normal people and that didn't bother me in that I realised that I had to do that to provoke things on a bit."*

In this experiment, 'computer usage' was measured by a single item concerning participants' use of a computer in their daily activities. Given the significant impact on of this explanatory variable on both presence and copresence, the final experiment on eye gaze and photorealism considered it in more detail by subdividing it into computer use, programming, experience with interactive virtual reality systems and computer games (see Section 5.2.2.10).

As mentioned in Section 2.5, limited empirical research has been conducted on the relationship between presence and copresence. The overall analysis indicates a significant positive correlation between reported presence and copresence, supporting Schroeder's argument that the two constructs are likely to interrelate [Sch02b]. However, it does not explain whether the relationship is causal, or whether it results from shared determinants. The fact that both presence and copresence were affected in a similar way by the introduction of the 'computer usage' variable in the analysis suggests that both may be at least partially shaped by expectations from prior media experiences. This is consistent with the findings of the grounded theory analysis reported in Chapter 6, where technical assumptions were found to have a significant impact on rational beliefs about the avatar's ability to play a significant role in the interaction. Overall, these findings point to the importance of exploring the impact of internal factors, or person variables, on social responses to virtual humans.

7.1.2 Responding to the agents as people

In terms of self-reported behaviour, computer usage was not found to have an effect. Participants in the *talking* condition were significantly more likely to want to interact with the agents. In terms of actual attempts to interact with the agents, all three animated conditions resulted in a significantly higher response than the *static* condition. On closer inspection, participants' desire or attempt to interact with the agents may not equate with seeing them as social entities, but with an almost gamelike motivation to engage with those aspects of the environment that appear to be most responsive. One question raised relates both to this experiment and to Bailen-

son et al.'s study on eye gaze [BBBL01], discussed in Section 2.4.4): how would participants' responses be affected by a responsive object, as opposed to a responsive humanoid agent?

More telling in terms of self-reported behaviour is the finding concerning the desire to avoid disturbing the virtual readers. One participant, who had the highest social anxiety score and who was in the *talking* condition, said that he made a deliberate effort to stay away from the table of virtual people throughout because he felt uncomfortable when they looked at him. He explained that this reaction was in no way different to how he would normally behave in a room full of real people. This comment supports the finding that higher social anxiety scores correlated significantly with a desire to avoid disturbing the agents.

7.1.3 Attributed sentience

As discussed in Section 2.5.6.2, Biocca et al.'s definition of social presence places particular emphasis on the notion of access to other minds: "Social presence is activated as soon as a user believes that an entity of the environment displays some minimal intelligence in its reactions to the environment and the user" [BBHS01, p.7]. The agents in the *responsive* and *talking* conditions in this experiment were designed to exhibit 'alert' behaviours in response to the participant's proximity. The findings support the hypothesis that these responsive agents would be perceived as being more 'aware.' In the post-experiment interviews participants confirmed that this sense of the agents being 'aware' was directly linked with their turning and gaze behaviours. This led one participant in the *responsive* condition to attribute a limited form of consciousness to them: *"You sense that they're aware of someone being there but maybe not aware in the way that a person with all senses and mobility would be, in that they didn't speak and they didn't look to interact. So it was a kind of muted awareness if you like. When they're the first to turn to you then I guess in some sense they did seem conscious."*

Although agents were considered to have higher 'awareness' in the conditions where they were visually responsive to the participant, this did not necessarily translate into a conscious attribution of sentience. Participants' comments illustrate that the illusion of 'minimal intelligence' was fragile, and not necessarily sufficient to sustain realistic social responses to the agents. For instance, participants explained that they did not maintain an appropriate interpersonal distance throughout the experience. One participant in the *moving* condition said:

> "At first I did feel aware of the people, I think I may have skated round them. So in that sense I suppose there was a real perception of them being there, a sort of confused social response. I knew that they were computer-generated, but I was fooled to a degree that there were these people there, so when I entered the space I

didn't barge straight into them. Eventually I thought 'this is a game' so I did barge into them and nothing happened."

This comment illustrates two themes echoed by several participants. The first concerns the temporal variations suggesting that copresence is an unstable response. The interviews support the idea voiced by Biocca et al. [BBHS01], Blascovich [Bla02] and Heeter [Hee01] that copresence, like presence, may fluctuate over time. Numerous participants mentioned that their reaction to the agents changed through the course of the experience. Typically, when something happened to break the illusion, such as a repeated animation, they began to react to them more as objects or as 'computer people.' Once the illusion broke, many participants reported trying to provoke a reaction by touching or walking through the agents. This marked a transition from respecting them as social entities to treating them as objects in a game.

The second theme concerns the tendency to respond on two levels. Many participants explained that the characters were 'not really' people. The sense that they were not real was compounded by their cartoonish appearance, and by any repetitive animation loops. In spite of this, several people mentioned their surprise that despite their rational realisation that the agents were computer-driven, they nonetheless responded to them on a separate level as people. In the words of one participant in the *moving* condition: *"I think my subconscious was aware that they weren't real, that I could mess about a bit. But my conscious was telling me how to react in that social situation. Just to begin with, I think. Then eventually I think my subconscious overrode that."*

As discussed in Section 2.5.6.4, Blascovich and colleagues' threshold model of social influence proposes two levels of response to virtual humans: a higher level that concerns rational reactions, and a lower level that concerns involuntary responses [BLB+01]. Comments by participants support this notion of a dual response to agents and avatars. They are also consistent with the finding from the grounded theory analysis (reported in Chapter 6) that suggested people were affected by the avatar's eye gaze in spite of any rationalisations about its inability to mirror their conversation partner's real movements.

7.1.4 Objective responses

Given the probable dual nature of responses to virtual humans, there is scope for studying subjective responses alongside objective involuntary responses. Bailenson et al. used tracking as a non-intrusive measure of proxemic behaviour [BBBL01]. This experiment used heart rate and EDA to compare objective responses to the training room and the virtual library containing the humanoid agents.

The findings for EDA illustrate a significant increase in mean levels from the training room to the virtual library room for all conditions except *moving*, which was just below the significance level. This may be attributable to the fact that the library environment offered a richer visual experience than the relatively bare training room. In terms of heart rate, however, only the *responsive* condition was associated with a significant increase in mean levels. This increase was found to diminish with increased computer usage, demonstrating a pattern consistent with the questionnaire findings for both presence and copresence.

This similarity between the heart rate and questionnaire data must be interpreted with caution because there is insufficient granularity to attribute psychophysiological changes directly to aspects of virtual human fidelity. Although individual events such as participants approaching the agents were marked in the data stream, these were not used in the analysis. This was primarily because the events were marked manually by an experimenter, and the potential margin of error was believed to be unacceptably high to draw any reliable conclusions. Also, the events were contingent on individual participants' actions in the VE. They were therefore unevenly distributed across the subject pool; given the highly individual nature of psychophysiological responses, it seemed there was insufficient foundation to explore the impact of individual events. Nevertheless, the similarity between the pattern of subjective copresence findings and objective heart rate findings is sufficiently encouraging to warrant further research into possible psychophysiological correlates for presence and copresence.

7.1.5 Summary discussion for experiment on responsiveness

Several themes emerged from the analysis. The regular use of computers was associated with a lower increase in heart rate between the virtual training room and the more 'social' library environment. It was also associated with a lower level of presence and copresence, lending support to the prediction from the grounded theory analysis that participant variables such as technical expertise can moderate the role of virtual humans in an interaction.

In terms of reported behaviour in response to the agents, the desire to avoid the agents correlated with participants' social anxiety scores, suggesting a tendency to behave in a way consistent with real-life social behaviour. As expected, the responsive agents were seen as being more aware, however this did not translate into an attribution of sentience to the agents. Participants suggested in the interviews that they responded to the agents on two levels, rationally considering them computer-generated characters, while sometimes responding to them on a separate level as social entities.

This section has discussed the findings from the experiment on responsiveness reported in Chapter 4. The following sections will discuss the findings from the two experiments on eye

gaze reported in Chapter 5.

7.2 Experiment on Eye Gaze

The first experiment investigating eye gaze, presented in Section 5.1, aimed to test whether having an avatar could improve perceived quality of communication as compared to not having a visual image at all. Along an imaginary continuum of mediated communication ranging from audio-only to full video, where would the avatar fall? More specifically, it aimed to explore the lower boundaries of behavioural realism by examining the impact of a single behaviour, eye gaze. Would a more realistic eye gaze model, inferred from conversational turntaking, outperform random gaze? Ultimately, answers to these questions could inform some recommendations on priorities for the improvement of communicative avatars.

The responses investigated included the degree to which the conversation was seen as being similar to a *face-to-face* conversation, the degree of *involvement* and *copresence*, and the degree of positive *partner evaluation*. The expectation was that in the context of dyadic interaction using this particular video-tunnel setup, the video condition would outperform the other conditions on all four measures. Since the task selected was an equivocal one in which participants would benefit from having visual feedback, it was predicted that video would result in better perceived communication quality since it presented participants with complete, real-time nonverbal feedback from the head and shoulders. It was therefore more visually informative than both the audio-only condition, which had no visual feedback, and the avatar conditions, which only used head, eye and lip movement.

It was not known whether having an avatar with random gaze would offer an improvement over not having a visual image at all. The overall analysis suggests that simply having an avatar whose head and eye movements are not related to the conversational flow does not improve communication when compared to audio-only. Indeed there is some evidence that in the case of copresence, the random-gaze avatar is worse than the pure audio stream. Comments in the interviews supported this. One participant in the *random-gaze* condition explained that: *"I didn't feel it represented anything. It just kind of sat there doing something rather than helping. It didn't make me feel there was a presence of the other person there."*

It was predicted that having an avatar whose gaze behaviour was directly related to the conversation would improve perceived quality of communication compared to one whose gaze behaviour was random. This prediction is supported by the results, which show that the inferred-gaze avatar consistently and significantly outperformed the random-gaze avatar. This builds on Colburn, Cohen and Drucker's findings [CCD00] that participants appeared to pay more at-

tention to the avatar when the gaze model was active than when it was not. This finding also supports Vilhjálmsson and Cassell's argument [VC98] that, for avatars to meaningfully contribute to communication, it is not sufficient for them to appear lively. Rather, their animation needs to reflect some aspect of the conversation that is taking place.

In the overall analysis the inferred-gaze avatar also significantly outperformed audio-only. This suggests that in contrast with a randomly animated avatar, one with meaningful animations can indeed contribute to perceived communication quality. In the case of the copresence variable, however, it performs worse than pure audio. The grounded theory analysis may partially explain this; participants explained that their sense of personal contact came primarily from the voice, and not from the avatar. The avatar's contribution was in other areas, for instance in helping to focus attention on the conversation.

What is perhaps most surprising is that the inferred-gaze avatar is not significantly different from the video in terms of sense of involvement and the extent to which the conversation is likened to face-to-face interaction. The avatar only used an approximation of a single nonverbal behaviour, eye gaze, whereas the video presented participants with full and accurate nonverbal feedback from the face. This is encouraging as it suggests that an avatar can begin to make a significant contribution to the positive perception of communication even without detailed facial expression.

The significant differences between the inferred-gaze avatar and the audio-only condition contrast with the subsequent findings of Bente and Kramer [BK02]. Comparing dyadic interaction using videoconferencing, audio-conferencing and avatars, they report no significant difference between audio-only and avatar-mediated interaction. A number of differences in experimental design may partly explain this. Firstly, their experiment concerned person perception, and therefore used different evaluative measures. Secondly, their avatar consisted of line drawings of the eyes, nose and hands, and was therefore significantly less detailed than those used here. It is impossible to speculate on the effect of the gaze animations, since there are no details on the model used. However, assuming that they were not tied to the conversation taking place, then the results are consistent with those reported here, namely that an avatar with random gaze does not offer an improvement over audio-only interaction.

7.2.1 Summary discussion for experiment on eye gaze

It should be stressed that this experiment was not designed to compare the relative merits of video and avatar-mediated communication. In this particular context it would undoubtedly be preferable to use video, as the results attest. Rather, this experiment was designed to explore whether an avatar with minimal behavioural fidelity could contribute to the experience of in-

teracting with another person. The grounded theory analysis presented in Chapter 6 identified several ways in which the avatar's contribution to communication was undermined by its failure to meet expectations. The fact that the interviews from this experiment highlighted significant shortcomings is not surprising given that the negotiation task placed high demands on the avatar's limited expressive potential. Nevertheless, the interviews were not able to address the potential differences between inferred and random gaze, since participants only experienced a single experimental condition. In this sense the grounded theory analysis and the analysis of the questionnaire responses are complementary.

The questionnaire findings focus on a direct comparison between the inferred and random-gaze avatars. Although both represent minimal behavioural fidelity, the inferred-gaze avatar was higher-fidelity in that its gaze animation was inferred from conversational turntaking. The questionnaire findings indicate that the random-gaze avatar did not offer a significant improvement over the audio-only interaction. However, the inferred-gaze avatar significantly outperformed both audio-only and the lower-fidelity random-gaze avatar on several response measures. This suggests that a minimal-fidelity avatar can begin to contribute to positive perceptions of the communication experience provided its behaviour reflects some aspect of the ongoing interaction. The inferred-gaze avatar did not, however, improve copresence. The purpose of the final experiment on eye gaze and photorealism, discussed in the following section, was to address unanswered questions from this experiment and to further investigate the impact of fidelity on copresence.

7.3 Experiment on Eye Gaze and Photorealism

The third and final experiment, reported in in Section 5.2, was designed to directly extend the first experiment on eye gaze. It therefore considered the same measures of perceived communication quality: *face-to-face*, *involvement*, *copresence* and *partner evaluation*. It also explored additional subjective responses, for instance a more detailed measure of copresence (called *social-copresence*) was created in light of analysis from the two previous studies. This measure consisted of three subcomponents: *general copresence*, *spatial copresence* and *personal contact*. Additional measures included *presence*, the degree to which avatar seen *as a person*, and its perceived *contribution* to conversation.

7.3.1 Perceived quality of communication

The findings in the first experiment were that the inferred-gaze avatar consistently outperformed the random-gaze avatar, and that for several of the response measures this difference was significant. However, the results confounded head tracking with inference about the avatar's eye

movement based on face-to-face dyadic research [AC76, AI72, Ken67]. The findings from this study resolved the ambiguity, since head-tracking was kept identical in all conditions. Independently of head tracking, the inferred-gaze model had a significant positive impact on perceptions of communication in the case of the higher-realism avatar.

The second aim was to compare gaze models within an immersive setting. Schroeder et al. point out that "Collaboration, particularly in highly immersive VEs, has not been extensively studied" [SSA+01, p.782]. Recall that previous studies by Fukayama et al. [FSO+01, FTM+02] and Lee, Badler and Badler [LBB02] were carried out in a non-immersive setting where the participants' point of view was controlled by the experimental setup. How would the eye gaze models perform in a communication context where participants were able to control their point of view within a shared 3D space? The results suggest that in the case of the higher-realism avatar, the pattern of results reported in the first experiment holds for three of the response variables: namely, that in the case of *face-to-face, involvement* and *partner evaluation*, the inferred-gaze model significantly outperforms the random-gaze model. This is consistent with the initial hypothesis that the inferred-gaze model should have a significant and positive impact on participants' responses to the communication experience in the IVE. The fact that this was not the case for the lower-realism avatar is noteworthy and is addressed below.

The third and final question concerned the appearance of the avatars. In the first experiment, both eye gaze conditions were implemented using the same relatively photorealistic male and female avatars. The present experiment aimed to investigate whether higher-quality avatar behaviour could compensate for a lower-realism appearance. It is clear that there is a highly consistent pattern of responses amongst the four subcomponents of quality of communication. This pattern also extends to the additional measures considered in Section 7.3.3 below. The overall conclusion must be that for the lower-realism avatar, the inferred-gaze model may not improve quality of communication, and may in some instances make things worse. However, for the higher-realism avatar, the inferred-gaze model improves perceived quality of communication. The evidence suggests that there should be some consistency between the type of avatar and the type of gaze model that is used: the more realistic the avatar's appearance, the more realistic the gaze model should be in order to contribute to the interaction experience.

7.3.2 The impact of person variables

The grounded theory analysis reported in Chapter 6 suggested that avatar fidelity does not work in isolation, but that personal characteristics and aspects of experimental design also affect responses. The current results support this. If *face-to-face* is taken as an example, participants' age was significant, with older participants more likely to 'suspend disbelief' and consider the

experience similar to a face-to-face interaction. Conversely, participants with a higher social anxiety score were less likely to rate it as similar to a face-to-face interaction. As in the first experiment, the people playing the role of the baker in the negotiation task had a lower *face-to-face* response. The fact that both studies found an effect for the role of the baker suggests that future studies might benefit from using a slightly modified task, for example changing the job descriptions in the scenario may ensure that both participants feel they have equal social status.

It is interesting that the type of interface used (Cave or HMD) had no effect. Schroeder, Widerström and colleagues' series of trials on a two-person collaborative puzzle task revealed that the type of interface used could significantly affect copresence [SSA+01, WAS+00]. Specifically, copresence was significantly reduced when the communication setup was asymmetrical with one immersed participant and the other using a desktop. In this experiment, however, the setup was relatively more symmetrical, with both participants immersed albeit using different interfaces. Another key difference is that Schroeder et al.'s task emphasised navigation and object manipulation, which are more likely to benefit from computational advantage. Conversely, the negotiation task used here was designed to assess the avatar's capacity to contribute to the conversation through nonverbal feedback and involved little direct engagement with the space. In fact, the VE was kept purposefully bare so as to minimise visual distraction from the avatar.

7.3.3 Additional measures

As mentioned above, this experiment considered additional measures. The two-item *copresence* measure from the first experiment was retained, in the interest of consistency. However, copresence was investigated in greater depth through additional variables that were thought of as possible subcomponents. These included *personal contact* and *spatial copresence*. The third was a *general copresence* measure that consisted of general descriptors such as 'being in the company of' and being 'together with' the conversation partner. The purpose of including these was to examine whether there is any benefit to discussing copresence in such general terms, as has often been the case in recent literature (see Section 2.5.4).

This new *social-copresence* measure was found to have a significant and positive correlation with presence. This is consistent with the findings from the experiment on responsiveness, which also revealed a significant correlation between presence and copresence. Recall, however, that the measure of copresence evolved during the course of the research. In the experiment on responsiveness, copresence was equated with a sense of personal contact and the sense of interacting with another person as opposed to with a computer interface. Both of these responses are also used as measures in the current experiment.

The same pattern of findings reported for perceived communication quality held for the new *social-copresence* variable, namely that there was a significant interaction effect between type of avatar and type of gaze. This was also highly significant for each of its subcomponents. The same interaction effect held again for additional measures including perceived avatar and gaze fidelity.

Both the first experiment and the later experiment reported by Lee, Badler and Badler [LBB02] used higher-realism humanoids. If the higher-realism avatar is considered in isolation, these results supports both the earlier findings concerning the benefit of inferred-gaze animations over random-gaze. However, this experiment extended the earlier work by also considering the impact of visual fidelity. Contrary to Fukayama et al. [FSO+01], it revealed a significant difference in the way the lower-realism and higher-realism avatars were affected by the different gaze models. The divergence in findings may be at least partially explained by two factors. Firstly, their gaze model was based on different parameters. Secondly, their communication context was fundamentally different, involving a one-way interaction from an agent to a human, in contrast to this experiment that involved a two-way communication between immersed human participants who were engaged in a delicate negotiation task. For this reason, it is likely that the demands placed on the virtual human were fundamentally different.

This experiment also extends the work of Bailenson et al. [BBBL01]. Their proxemics experiment, discussed in Section 2.4.4, considered both objective tracking data and a 5-item subjective measure of social presence. Their findings indicated that increased gaze fidelity correlated with higher social presence, whereas increased visual fidelity had no effect. In terms of appearance, their lower-realism agent had a lower polygon count, making its facial mesh appear more artificial than the otherwise identical higher-realism agent. Nevertheless, the difference in appearance was marginal, whereas the gender-neutral, lower-realism avatar in this experiment was significantly less detailed than the higher-realism male and female avatars.

Despite its minimal visual and behavioural fidelity, even the lower-realism avatar seemed to contribute to the interaction. In the words of of one participant in the lower-realism, random-gaze condition: *"Even if it is not a very realistic avatar, it helps a little. It gives you something to focus on. Although you do not think of it as a person, strangely it does stop you turning away or doing anything inappropriate. Also your mind does not wander as much as it might on the telephone. You are immersed in the environment."* Many participants mentioned that the avatar helped to give them a strong sense of being in a shared space with their partner. Without exception, all participants stood facing their partner's avatar throughout the entire conversation. They took care to maintain a suitable interpersonal distance and felt compelled to display polite

attention.

7.3.4 Summary discussion for experiment on eye gaze and photorealism

Previous, non-immersive studies using higher-realism virtual humans indicated that inferred-gaze animations outperformed random-gaze animations on a number of subjective response measures. This experiment extended previous research to consider the performance of the gaze animations in relation to visual fidelity. The findings revealed an overwhelmingly cohesive model indicating a significant interaction effect between type of avatar and type of gaze. This has significant new implications for avatar design, because it suggests that the impact of gaze animations can change depending on the appearance of the avatar.

7.4 Chapter Summary

This chapter has discussed the findings from the three experiments presented in Chapters 4 and 5. The experiments considered three aspects of fidelity: responsiveness, eye gaze realism, and photorealism. In substantive terms, the overarching question addressed was whether minimal increases in virtual human fidelity could impact on experiences of social interaction in CVEs. In methodological terms, the goal was to explore ways of measuring social responses including copresence. The measure of copresence used in the third experiment was more detailed than in the first, exploring possible subcomponents highlighted by previous analysis. Although the measures evolved over the course of the research, the consistency in specific key findings suggests the results should be viewed incrementally rather than as disconnected from each other. For instance, presence and copresence were found to be significantly correlated in both the experiment on responsiveness (Chapter 4) and on eye gaze and photorealism (Section 5.2). In addition, the results for perceived communication quality were very similar for the two eye gaze experiments where the higher-realism avatar is concerned.

In summary, the three experiments presented in this thesis investigated the impact of three key aspects of fidelity investigated on a variety of social responses. The results suggest that an avatar with minimal fidelity can indeed contribute to positive perceptions of the interaction experience, and that even minimal behavioural fidelity can present an improvement over mere liveliness. The findings also indicate a strong interaction effect between photorealism and eye gaze behaviour, suggesting a relationship between visual and behavioural fidelity that warrants further exploration. In addition to avatar fidelity, the findings also suggest that the context of interaction as well as the individual characteristics of different users are likely to shape responses to avatars. The following chapter will summarise the overall contributions of this research and suggest directions for future research.

Chapter 8

Conclusions

At present, CVEs are primarily used for research and entertainment purposes and have yet to come into mainstream use as a communications medium. They have the potential, however, to extend beyond their present usage to address the practical collaborative needs of geographically dispersed users. One of their chief attractions lies in their ability to combine 3D spatial interaction with a high degree of multisensory immersion. They are therefore of particular interest for those collaborative situations, such as remote acting rehearsals, where it is essential to preserve spatial relationships among users.

One key barrier to effective communication in current CVEs is the relative paucity of avatar expressiveness as compared to live video. However, increasing the expressive potential of avatars involves significant challenges. In terms of their appearance, the tension between realism and real time means that photorealism comes at the expense of unwanted delays to real-time communication. Visual fidelity must therefore be traded off against available computing resources. In terms of behaviour, the tension between control and cognitive load underlines the difficulty of transparently driving avatar behaviours that appropriately represent the user. Full manual control of avatar behaviour would entail an unacceptable level of cognitive load. On the other hand, full tracking can be expensive and invasive, and may not be desirable in a medium that is prized for the control it affords over visual identity.

Given these constraints, the approach taken in this research was to explore the lower boundaries of virtual human fidelity. The overarching goal was to investigate whether minimal increments in fidelity could contribute to participants' interaction experience. Fidelity was considered in terms of both dynamic behaviour (behavioural fidelity) and static appearance (visual fidelity).

This research focused primarily on subjective measures, combining questionnaires with an in-depth qualitative analysis of interviews with participants. It also explored the use of objective psychophysiological responses. Three experiments were conducted investigating three key

aspects of fidelity: behavioural fidelity in terms of responsiveness and eye gaze behaviour, and visual fidelity in terms of photorealism.

8.0.1 Experiment on responsiveness

The experiment on responsiveness reported in Chapter 4 compared people's responses to visually identical agents whose behaviours were modified to reflect increasing levels of responsiveness. The goal was to isolate the impact of nonverbal behaviours from verbal interaction to understand how simple increments in animation and basic responsiveness could contribute to the sense of copresence. Other responses of interest included the possible attribution of sentience to the agents, and self-reported behaviour in response to them. The social context was deliberately unstructured, with participants left free to explore a virtual library containing a group of seated readers.

The findings indicate that increasing avatar responsiveness even on a simple level can have a significant impact on certain aspects of people's social responses. The results indicated that participants who encountered the static agents experienced a significantly lower sense of copresence than those who encountered animate or responsive agents. However, when copresence was considered in relation to participants' level of everyday computer use, then the visually responsive agents resulted in a significantly higher response. Also, more experienced computer users were less likely to experience copresence. Interestingly, this pattern was precisely reflected in the heart rate data, suggesting a parallel between the subjective and objective responses.

None of the participants reported responding to the avatars as if they were 'really' people. However, many expressed surprise at the fact that they had respected some social norms despite the fact that they knew the avatars were computer-generated. People with higher levels of social anxiety were significantly more likely to avoid disturbing the avatars, supporting previous findings that on some level people can respond to avatars as social actors even in the absence of two-way verbal interaction. The interviews illustrated that there were a complex variety of factors at play, including the behaviours of the avatars, people's prior experience with technology and their responses to real-life social situations.

8.0.2 Experiment on eye gaze

The first experiment on eye gaze, reported in Section 5.1, sought to answer two questions. First, whether an avatar with minimal behaviour can improve the quality of communication between two remote users, or whether it simply acts as a placeholder or distraction. Second, whether an avatar whose behaviour is directly related to the conversation can offer a significant improvement over a visually identical avatar with random behaviour.

The experiment was deliberately conducted using a non-immersive videotunnel setup in order to isolate gaze behaviour from any other factors, such as spatial, gestural or postural cues that might have confounded results. Participants saw a face-on, head-and-shoulders view of the avatar representing their conversation partner. In the inferred-gaze condition, the avatar's head movement was tracked and its eye movement was driven by the audio stream based on 'while speaking' and 'while listening' animations whose timings were taken from research on face-to-face dyadic interaction. In the random-gaze condition, the participant's head was not tracked, and both the avatar's head and eye movement were random. The results showed the inferred-gaze avatar significantly outperformed the random-gaze one on several response measures. It also significantly outperformed the audio-only condition.

The random-gaze avatar did not provide a significant improvement over pure audio, suggesting that the simple introduction of an avatar does not automatically improve participants' perception of communication. Rather, the avatar must have certain behaviour characteristics in order to be useful. The inferred-gaze avatar outperformed the pure audio stream on several measures. This suggests that an avatar whose behaviours reflect an aspect of conversational flow can indeed make a contribution to improving remote communication. Finally, the inferred-gaze avatar significantly outperformed the random-gaze avatar on all measures, indicating that an avatar whose behaviours are related to the conversation can present a marked improvement over an avatar that merely exhibits liveliness.

8.0.3 Experiment on eye gaze and photorealism

The experiment presented in Section 5.2 sought to extend the work of the previous studies by investigating the impact of both visual and behavioural fidelity on perceived quality of communication between participants meeting in a shared IVE. In terms of appearance, the avatar was either visually simplistic or more realistic; in terms of behaviour, the random-gaze and inferred-gaze models from the previous experiment were tested in this more demanding immersive setting.

The findings cleared up an ambiguity from the previous study regarding whether the significant differences in performance between the gaze models were due to head-tracking or to eye animations inferred from the audio stream. The conclusion was that independent of head-tracking, inferred eye animations can have a significant positive effect on participants' responses to an immersive interaction. The caveat is that they must have a certain degree of visual realism, since the lower-realism avatar did not appear to benefit from the inferred-gaze model. This finding has implications for inexpensive ways of improving avatar expressiveness using information readily available in the audio stream. It suggests avenues for interim solutions for the

difficult problem of providing robust eyetracking in a Cave.

8.1 Key Questions Addressed in the Research

Chapter 1 introduced three central research questions concerning the impact of avatar fidelity:

1. *Can an avatar contribute to positive perceptions of the interaction experience?*
2. *Can minimal behavioural fidelity offer an improvement over mere 'liveliness'?*
3. *What is the relationship between visual and behavioural fidelity?*

The first question addresses the underlying premise of whether it is at all possible for an avatar that is neither photorealistic nor literal in its behaviour to contribute to the experience of interacting in a CVE. The question that follows immediately from this is whether an avatar with minimal behavioural fidelity can offer a meaningful improvement over one whose behaviour is in no way tied to the ongoing conversation. Given the need for tradeoffs, a strategy adopted by several researchers has been to prize behavioural fidelity over visual fidelity in the development of communicative avatars. The third question focuses on the interrelationship between these two aspects of fidelity.

The overarching goal of the research was to investigate whether minimal fidelity can begin to give users a sense of 'being together' in the CVE even though they know rationally that the virtual space and the virtual humans in it are not 'real'. This sense of being together is referred to as copresence. Chapter 2 discussed in detail the lack of consensus on its definition and measurement. Therefore a fourth related question concerned the assessment of people's experiences:

4. *How should copresence be defined and measured?*

8.1.1 Main contributions

Research into avatar-mediated communication is arguably still in its infancy, and as rendering and animation techniques continue to mature there will likely be numerous improvements in terms of visual and behavioural fidelity. In terms of assessing people's interaction experiences in CVEs, views on the definition and measurement of key constructs such as copresence are continually developing. At present, however, there is scope for exploring the impact of different aspects of avatar fidelity with a view to understanding how minimum cues can achieve maximum results in terms of people's interaction experiences in CVEs. This thesis made both substantive and methodological contributions. The substantive contributions consist of empirical findings concerning the impact of specific aspects of avatar fidelity on a range of responses.

The methodological contributions concern proposed research approaches to studying these relatively new phenomena.

8.1.1.1 Methodological contributions

The critical literature review presented in Chapter 2 discussed the problematic use of the terms *social presence* and *copresence* in the recent literature on virtual humans and CVEs. This thesis used the term copresence to mean the aggregate of social responses contributing to the sense of being with another person in the CVE. One of the methodological contributions of this thesis was the development and gradual refinement of questionnaire items relating to a selection of social responses including the sense of personal contact, spatial copresence, attributed sentience, and perceived contribution of avatars to the interaction process.

These measures evolved through the course of the research. Their development was greatly aided by the use of qualitative analysis of the interview responses from the first experiment on eye gaze. For example, the grounded theory analysis illustrated that two potential subcomponents of copresence (personal contact and spatial copresence) may have distinct root causes and therefore need to be considered individually.

Grounded theory is an iterative process whereby analysis shapes successive sampling decisions. Sampling is usually restricted to a relatively small number of respondents carefully chosen in light of ongoing analysis to explore a broad range of experiences relating to the phenomenon of interest. The research in this thesis departs from standard procedure by applying the method to a comparatively large number of participant interviews concerning a single, focused experience of avatar-mediated interaction. Additional methodological contributions therefore include suggestions on techniques for visual abstraction from verbal data (Chapter 6) and for validating the analysis (Chapter 3).

8.1.1.2 Substantive contributions

The main substantive contributions directly address the three central questions concerning avatar fidelity posed at the beginning of the thesis. The first question asked whether an avatar with minimal fidelity could contribute to interaction experience. The findings from the first experiment on eye gaze indicate that it can, provided that its gaze behaviour reflects some aspect of the ongoing conversation, in this case something as simple as turntaking. These findings also address the second question, which asked whether minimal increases in behavioural fidelity offer an improvement over mere 'liveliness'.

The third question concerned the relationship between visual and behavioural fidelity. The discovery of a significant and overwhelmingly consistent interaction effect between visual and behavioural realism constitutes an important contribution of this thesis. The findings from the

experiment on eye gaze and photorealism indicated that the higher-realism avatar benefited from the higher-fidelity eye gaze animations, whereas the opposite was true for the lower-realism avatar. This suggests the need to align behavioural fidelity with avatar appearance. The conclusion is that the impact of behaviour is not independent of appearance and points to a more complex picture than was previously envisaged. Simply privileging behavioural over visual fidelity may not lead to optimal improvements for expressive avatars.

Additional findings indicate that avatar fidelity does not work in isolation in shaping interaction experiences. Both communication context and personal characteristics such as everyday social anxiety, prior media experiences, and technical expertise shape perceptions of the avatar's role in interaction. Interviews with participants indicated that it is possible to rationally think of avatars as computer-generated and therefore not 'real', while simultaneously exhibiting social responses towards them. This is consistent with Reeves and Nass' theory of the medium as social actor, that predicts people will tend to anthropomorphise media and treat them as social entities [RN96]. This finding also supports Blascovich et al.'s hypothesis that there are at least two levels of response to virtual humans: higher-level rational responses, and lower-level involuntary responses [Bla02].

Presence was not the main focus of the thesis. It was nevertheless measured in both the experiments where participants experienced the VE immersively. Chapter 2 mentioned the need for empirical investigation into the relationship between presence and copresence. A significant positive correlation between presence and copresence was found in both the experiment on responsiveness and the one on gaze and photorealism. This finding contributes additional support for the possible relationship between these constructs, although it does not of course explain the nature of the relationship and whether it is causal or merely a function of shared determinants.

The grounded theory analysis contributed a finding that has implications for different approaches to the difficult problem of driving an avatar's behaviours in real time. Discussions of 3D graphical chats have emphasised the attraction of safe identity play in an environment where users can control their appearance and actions. This analysis suggests that the ability to control behaviour may heighten the association between the avatar and the person represented, while undermining trust. The concern is that the avatar allows for the masking of real emotions, so that it does not necessarily give an honest measure of the person one is interacting with. On the other hand, the accurate portrayal of a person's real actions through tracking is likely to result in a high degree of association as well as trust. However, users are likely to need a guarantee that the avatar accurately mirrors the real-life actions of their conversation partners.

Between tracking and manual control is the option of partially automating a selection

of behaviours. This was the approach taken in animating the inferred-gaze avatar in the two experiments on eye gaze. The quantitative findings suggests that simply adding 'liveliness' to the avatar's behavioural repertoire does not add value to the interaction. The qualitative findings further suggest that users would benefit from information about how 'truthful' the animations are, because in the absence of priming people are likely to interpret the significance of the avatar's behaviour according to their own technical assumptions about how it is driven. These assumptions are sometimes illogical and uninformed, and may work to the detriment of the avatar by leading users to discard as insignificant even those selected behaviours that are in fact informative. Overall this analysis signals the importance of educating users about the behavioural capabilities of the avatar.

These findings are based on three laboratory-based experiments, two of which used high-end immersive equipment. Reservations about the ecological validity of these results must therefore be tempered by the consideration that immersive technology is not in mainstream use. Desktop use is limited primarily to social interaction in online graphical chat environments. It is therefore difficult to predict how these findings would generalise to a real-world setting, particularly since they consider one-off interactions rather than long-term use. This was done purposefully to maximise the demands placed on the avatars. The two experiments on eye gaze tested how the avatars would perform in a communication context where demands for nonverbal feedback were high, but no allowances were made for their failings based on habituation or knowledge about their functionality.

The conflict negotiation task developed for the eye gaze experiments meant that the avatars were assessed in a deliberately unforgiving context. As discussed in Section 5.1.2.6, it is believed that the demand for visual feedback is higher in equivocal situations that have no single 'correct' outcome but require negotiation. It was also thought that the emotional content of the scenario combined with the negotiation requirements of the task would mean that results could speak both to social and business contexts. The task resulted in lively conversations that greatly contrasted with the stilted exchanges in the pilots using a moral dilemma task. It was therefore successful from the point of view of providing a sufficiently rich topic of conversation for two people who did not know each other. The fact that the role played had an effect on certain responses suggests, however, that the scenario could benefit from slight modifications if used in future studies.

8.2 Directions for Future Work

The research focused primarily on subjective responses, based on the rationale that higher-level conscious responses play a central role in shaping users' media choices. As argued by Short, Williams and Christie [SWC76], subjective perceptions are particularly relevant to the potential adoption or rejection of an emerging communications medium.

The use of questionnaires is common in presence research. However, post-test questionnaires obviously do not tell the whole story; they fail to capture any temporal variations in responses during the course of the experience, and are also prone to memory effects. The added problem is that the research agenda is dictated by the preconceptions of the researcher. Avatar-mediated interaction represents a relatively new research area that can benefit from qualitative research. Qualitative methods focus on those aspects of experience that are considered important by potential users, as opposed to the creators of the system.

The grounded theory analysis detailed the connections between various key factors in participants' interaction experience, from prior expectations to aspects of the communication context, to properties of the avatar's appearance and behaviour and finally to communication outcomes. This research area could benefit from further use of qualitative methods to obtain a richer understanding of participants' views on interaction in CVEs than is possible through anecdotal reports of general themes from interviews. The use of grounded theory implies a significant time commitment, but is beneficial for hypothesis generation and for identifying potentially fruitful research directions.

Although interviews and in-depth qualitative analysis go a step further than questionnaires in understanding users' perceptions, they necessarily only capture the rational level of experience. The findings reported in this thesis indicate that it is important to further explore both higher-level and lower-level social responses to virtual humans. Previous research has indicated that minimal fidelity can affect lower-level involuntary responses such as spatial behaviour in response to an agent [BBBL01]. Potentially fruitful avenues for research include the observation of involuntary behaviours, as well as the use of psychophysiological measures to study objective responses.

The argument for investigating psychophysiological responses such as heart rate and EDA is that they can capture the experience itself, as opposed to post-hoc rationalisations. They have the potential to highlight temporal variations and aspects of the experience that enhance or hider presence and copresence. The use of psychophysiological measures in the experiment on responsiveness was only exploratory, and not the primary focus of the research. It is nevertheless interesting that the findings for heart rate followed the same pattern as those for subjective

reports on presence and copresence; the responsive agents resulted in a higher response than those that were static or simply moving, but this response was tempered by increased computer use.

Biocca, Harms and Burgoon point out that there is as yet no firm empirical connection between subjective and objective measures: "We are unaware of any explicit measures of mediated social presence. This may be due to the absence of any consistent psychophysiological signature simply for the presence of another even though interaction with others may elicit various psychophysiological responses depending on context and interaction ... Social presence definitions would need to be more concrete to allow for valid physiological measures" [BHB02, p.19]. Although the present finding does not establish a direct connection that could warrant replacing subjective post-test measures with objective continuous measures, the analysis suggests that psychophysiological correlates for social responses would be a worthwhile avenue for future research. The exploration of objective approaches should, however, proceed in tandem with the further development of subjective approaches, because the definition of copresence itself needs further refinement.

In both the experiments presented in Chapter 5, eye gaze was taken as a specific (though important) instance of avatar behaviour. One cannot claim, of course, that the results will generalise to other aspects of avatar behaviour, but findings for eye-gaze will generate hypotheses for studies of further aspects of avatar animation. Future research should aim to expand on these findings by investigating the impact of additional behaviours such as facial expression, gesture and posture. The gaze studies focused on dyadic interaction. However, gaze is an essential cue in small-group conversation management, and future studies could therefore expand the interaction context to include multi-party groups of 3 or more.

The experiments on eye gaze attempted to focus on the impact of a single behaviour. The isolation of a single behaviour is potentially problematic, however, because it compromises the gestalt of behaviours experienced in face-to-face interaction. Short, Williams and Christie point out the danger of treating behaviours individually: "In attempting to assess the functions of the visual channel, it is dangerous to confine attention at any one time to individual cues such as posture, eye-gaze, proximity and the like. The channels do interact... Studies of media must look at relevant combinations of channels. Important overall properties of communication may be missed if attention is restricted to individual channels" [SWC76, p.57-8]. Additional behaviours should therefore be investigated in conjunction with gaze and with each other with a view to exploring their interdependencies.

Another important direction for future research concerns self-representation. The research

presented in this thesis deliberately prevented participants from seeing their own avatar in full in order to avoid clouding perceptions of the avatar representing the conversation partner. It therefore did not investigate their relationship to their own avatar, which is also likely to be a central factor in avatar-mediated interaction. The degree of visual and behavioural fidelity of users' own avatars may influence the degree to which other avatars are perceived to be a viable representation of other users. The impact of self-presentation on the significance of others' avatars is an open question for empirical validation.

It is possible to imagine a day when avatars will, like those in the novel *Snow Crash* [Ste92], communicate users' intentions so reliably that people will willingly use CVEs for social interaction and for serious collaborative purposes. Today's avatars display an extremely limited expressive potential compared to those depicted in cyberfiction and recent films. They are restricted in their visual and behavioural fidelity by nontrivial technical constraints. In 1998 Allbeck and Badler [AB98] argued that every aspect of avatar design, motion and appearance described in *Snow Crash* [Ste92] had already begun to be tackled by different research groups. It is encouraging to think that all of these various branches of research may eventually converge to create truly compelling communicative avatars.

This thesis has aimed to investigate the impact of minimal avatar fidelity on perceptions of interaction experience in CVEs. Research was presented investigating key aspects of visual and behavioural realism. The findings suggest that a cartoonish avatar with minimal behaviours can begin to positively affect perceptions of interaction. The caveat is that its behaviour should reflect some aspect of the ongoing interaction, and the level of behavioural realism should be aligned with the photorealism of the avatar's appearance. Future work will build on these findings by combining subjective and objective approaches to understand how the sense of copresence with avatars can be enhanced for richer multiparty interaction in CVEs.

Appendix A

Summary of the three experiments

Table A.1: Experimental design and running

Focus of experiment	Responsiveness	Eye gaze	Eye gaze + Photorealism
Logical order	Chapter 4	Section 5.1	Section 5.2
Chronological order	2	1	3
Where conducted	UCL	BT Exact	UCL
Setup	Immersive (Cave)	Non-immersive (videotunnel)	Immersive (Cave + HMD)
Task	Exploration of VE	Conversation (negotiation)	Conversation (negotiation)
Number of participants	41	100	48
Virtual humans	Agents	Avatar	Avatar
Visual fidelity	• Semi-photorealistic	• Semi-photorealistic	• Semi-photorealistic • Cartoonish
Behavioural fidelity	• Pre-scripted general behaviours in relation to participant's proximity	• Inferred-gaze • Random-gaze	• Inferred-gaze • Random-gaze
Conditions	1. Static 2. Animate 3. Responsive 4. Talking	1. Video 2. Inferred-gaze avatar 3. Random-gaze avatar 4. Audio-only	1. Lower realism + Random gaze 2. Lower realism + Inferred gaze 3. Higher realism + Random gaze 4. Higher realism + Inferred gaze

Table A.2: Analysis and main findings

Focus of experiment	Responsiveness	Eye gaze	Eye gaze + Photorealism
Responses analysed	• Questionnaire • Interview • Psychophysiological measures (Heart rate and EDA)	• Questionnaire • Interview (Analysed using grounded theory)	• Questionnaire • Interview
Focus of post-test questionnaire	1. Copresence 2. Presence 3. Participant behaviour 4. Perceived agent awareness	1. Face-to-face 2. Involvement 3. Copresence 4. Partner evaluation	1. Face-to-face 2. Involvement 3. Copresence 4. Partner evaluation 5. Social-copresence 6. Avatar fidelity 7. Gaze fidelity
Main findings	Significant interaction effect between condition and participants' computer experience in terms of presence, copresence and heart rate. Higher agent responsiveness associated with greater attributed sentience.	Video Significantly outperforms other conditions on all measures. Inferred-gaze outperforms random-gaze avatar on all measures, and outperforms audio-only on all measures except for copresence.	Significant interaction effect between visual and behavioural fidelity. Higher-realism, inferred-gaze animation positively affects higher- perceptions of higher-realism avatar but not lower-realism avatar.
Publication	[GSPR03]	[GSBS01]	[GSV+03]

Appendix B

Materials for the Experiment on Responsiveness

This appendix contains the materials used for the experiment on responsiveness, reported in Chapter 4. These include, in order of appearance:

1. The instruction sheet given to participants at the start of each session.
2. The consent form.
3. The adjusted Social Avoidance and Distress questionnaire [WF69] (questions added to detract participants' attention from the social focus of the questionnaire appear here in italics. Answers to these questions were not used in the analysis).
4. The post-test questionnaire administered at the end of each session.
5. The semi-structured interview agenda.

Information Sheet for Participants

Thank you for volunteering to participate in this study. This is one in a long series of studies aimed at understanding people's responses to virtual environments.

Please read through this information and feel free to ask any questions. While the experimenters will be happy to answer any general questions you may have, they have been instructed not to discuss some aspects of the study until the end.

This study will take place inside a virtual reality room called a CAVE. You will be asked to wear a small tracking device attached to the head with a small adjustable strap. You will be asked to put on a pair of lightweight stereo glasses; these can be worn over eyeglasses if necessary. You will also be asked to take off your shoes to avoid marking the floor of the CAVE.

Procedures

- You will be asked to sign a Consent Form
- You will be fitted with sensors to measure your heart rate and galvanic skin response (palmar sweating)
- You will then be led through to the CAVE
- You will have a brief practice period to get used to moving around the virtual environment in the CAVE
- Once you are comfortable, your task will be explained to you. You will be asked to move into another part of the virtual environment and observe it for 3-4 minutes
- Afterwards you will be asked to complete some further questionnaires about your experiences
- Finally there will be a discussion with the experimenters

The whole study is expected to take less than 1 hour.

Please note

- All your questionnaire responses will remain entirely confidential
- No identifying information about participants will be published in any form
- Please do not discuss this study with anyone for ONE MONTH. This is because the study is continuing and you may happen to speak to someone who may be taking part.

- You are free to withdraw from the study *at any time* and *without giving reasons for withdrawing*.

A note about virtual reality equipment:

When people use virtual reality systems, some people sometimes experience some degree of nausea. If *at any time* you wish to stop taking part in the study due to this or any other reason, please say so and we will stop.

There has been some research which suggests that people using virtual reality displays might experience some disturbances in vision afterwards. No long-term studies are known to us, but the studies which have been carried out do testing after about 30 minutes, and find the effect is still sometimes there.

There have been various reported side effects of using virtual reality equipment, such as "flashbacks."

With any type of video equipment there is a possibility that an epileptic episode may be generated. This, for example, has been reported for computer video games. For this reason we regret that we are unable to accept volunteers who are known to have suffered from epilepsy.

Virtual Reality Study Consent Form

Please read and answer the following questions carefully:

Have you read the information sheet about this study?	YES/NO
Have you had an opportunity to ask questions about the procedure?	YES/NO
Have you received satisfactory answers to all your questions?	YES/NO
Have you received enough information about this study?	YES/NO
Do you understand that you are free to withdraw from this study *at any time* and *without giving a reason for withdrawing*?	YES/NO
Do you understand and accept the risks associated with the use of virtual reality equipment?	YES/NO
Do you agree to take part in this study?	YES/NO

We would like to videotape you when you are in the virtual environment, and would also like to audiotape the conversation at the end. These tapes will be used for data analysis purposes only and will be kept entirely confidential.

Do you agree to be videotaped?	YES/NO
Do you agree to be audiotaped?	YES/NO

Please check:

I certify that I do not have epilepsy

I certify that I will not be driving a car, motorcycle, bicycle, or use other types of complex machinery that could be a danger to myself or others, within 3 hours after the termination of the study

Signed..... Date.....

Name in block letters.....

This questionnaire is composed of 28 statements regarding your feelings in social gatherings. Circle YES if you consider that the statement if true of your feelings most of the time. Circle NO if you consider that the statement is rarely true of you. Remember that this information is completely *confidential.*

	Please circle	
I am not easily distracted when I am working	YES	NO
I would prefer to work in an open plan office	YES	NO
I find it difficult to concentrate if there is a lot of noise	YES	NO
I normally prefer factual books to fiction	YES	NO
I normally ensure I have adequate lighting before I start reading	YES	NO
I rarely read books for pleasure	YES	NO
The amount of natural light in my workplace is important to me	YES	NO
I concentrate better at night than during the day	YES	NO
I often read books that my friends have recommended	YES	NO
I enjoy reading	YES	NO
I rarely misplace important personal items	YES	NO
I find I need to take frequent breaks in order to work effectively	YES	NO
My physical surroundings at work are not important to me	YES	NO
I would avoid taking a job that involved working through the night	YES	NO
I frequently look for opportunities to take on more responsibility at work	YES	NO
I consider myself to be a tidy and orderly person	YES	NO
I am often late for appointments or meetings	YES	NO
I feel relaxed even in unfamiliar social situations	YES	NO
I try to avoid situations which force me to be very sociable	YES	NO
It's easy for me to relax when I am with strangers	YES	NO
I have no particular desire to avoid people	YES	NO
I often find social settings upsetting	YES	NO
I usually feel calm and comfortable in social situations	YES	NO
I am usually at ease when talking to someone of the opposite sex	YES	NO
I try to avoid talking to people unless I know them well	YES	NO
If the chance comes to meet new people, I often take it	YES	NO
I often feel nervous or tense in casual get-togethers in which both sexes are present	YES	NO
I am usually nervous with people unless I know them well	YES	NO
I usually feel relaxed when I am with a group of people	YES	NO
I often want to get away from people	YES	NO
I usually feel uncomfortable when I am in a group of people I don't know	YES	NO
I usually feel relaxed when I meet someone for the first time	YES	NO
Being introduced to people makes me tense and nervous	YES	NO
Even though a room is full of strangers I may enter it anyway	YES	NO
I would avoid walking up to and joining a large group of people	YES	NO
When my superiors want to talk to me, I talk willingly	YES	NO
I often feel on the edge when I talk to a group of people	YES	NO
I tend to withdraw from people	YES	NO
I don't mind talking to people at parties or social gatherings	YES	NO
I am seldom at ease in a large group of people	YES	NO
I often think up excuses in order to avoid social engagements	YES	NO
I try to avoid formal social occasions	YES	NO
I usually go to whatever social engagements I have	YES	NO
I find it easy to relax with other people	YES	NO

Virtual Reality Study Questionnaire

Please note that in this questionnaire *"laboratory"* refers to the physical space in which the study is taking place, and *"room"* refers to the virtual room you explored.

1. Please rate the extent to which you were aware of background sounds in the laboratory in which this experience was actually taking place. Rate this on the following scale from 1 to 7 (where for example 1 means that you were not at all aware of the background sounds):

During the experience I was aware of background sounds from the laboratory ...								
Not at all	1	2	3	4	5	6	7	Very much so

2. How dizzy, sick or nauseous did you feel resulting from the experience, if at all? Please answer on the following 1 to 7 scale.

I felt sick or dizzy or nauseous during or as a result of the experience...								
Not at all	1	2	3	4	5	6	7	Very much so

3. Gender:

1. Male	1
2. Female	2

4. Please rate your *sense of being in the* room, on the following scale from 1 to 7, where 7 represents your *normal experience of being in a place.*

I had a sense of being there in the room...								
Not at all	1	2	3	4	5	6	7	Very much so

5. To what extent were there times during the experience when the room was the reality for you and you almost forgot about the real world of the laboratory where the experience was really taking place?

There were times during the experience when the room was the reality for me...								
At no time	1	2	3	4	5	6	7	Almost all of the time

6.

My status is as follows:	*Please tick against your answer*
1. Undergraduate student	1
2. Masters student	2
3. PhD student	3
4. Research Assistant/Research Fellow	4
5. Staff member - systems/technical staff	5
6. Faculty	6
7. Administrative staff	7
8 Other (please write in)...	8

7. When you think back about your experience, do you think of the room more as *images that you saw*, or more as *somewhere that you visited*?

The room seems to me to be more like...								
Images that I saw	1	2	3	4	5	6	7	Somewhere I visited

8. Have you experienced 'virtual reality' before?

I have experienced virtual reality...								
Never before	1	2	3	4	5	6	7	A great deal

9. During the course of the experience, which was strongest on the whole, your sense of being in the room, or of being in the real world of the laboratory?

I had a stronger sense of...								
Being in the lab	1	2	3	4	5	6	7	Being in the room

10. Overall, how well do you think that you achieved your task?

I achieved my task...								
Not at all	1	2	3	4	5	6	7	Fully

11. To what extent do you use a computer in your daily activities?

I use a computer...								
Not at all	1	2	3	4	5	6	7	Almost all the time

12. During the time of the experience, did you often think to yourself that you were just standing in a laboratory or did the room overwhelm you?

During the experience I was thinking that I was really in laboratory...								
Most of the time	1	2	3	4	5	6	7	Rarely

13. During the course of the experience, how much were you aware of the experimenters?

Not at all	1	2	3	4	5	6	7	Very much

14. During the course of the experience, did you have a sense that you were in the room with other people or did you have a sense of being alone?

With other people	1	2	3	4	5	6	7	Alone

15. How aware were you of the characters in the room?

Not at all	1	2	3	4	5	6	7	Very much

16. How closely did you observe the characters?

Not at all closely	1	2	3	4	5	6	7	Very closely

17. How curious were you about the characters?

Not at all curious	1	2	3	4	5	6	7	Very curious

18. When you *first* saw the characters, to what extent did you respond to them as if they were people?

Not at all	1	2	3	4	5	6	7	Very much

19. Now consider your response *over the course of the whole experience*. To what extent did you respond to them as if they were people?

Not at all	1	2	3	4	5	6	7	Very much

20. Were the characters moving?

☐ Yes ☐ No ☐ I don't know

21. Did the characters interact with you in any of the following ways (please check "yes" or "no" for each):

	Yes	No
Smile	☐	☐
Frown	☐	☐
Make eye contact	☐	☐
Wave	☐	☐
Talk	☐	☐
Turn to face you	☐	☐

22. How did the characters seem to respond to you? Did they seem:

☐ Extremely unfriendly

☐ Unfriendly

☐ Neither friendly nor unfriendly

☐ Friendly

☐ Extremely friendly

23. Did the characters seem like they:

☐ Knew each other ☐ Didn't know each other ☐ I don't know ☐ I didn't notice

24. How much did the characters seem to respond to you?

Not at all	1	2	3	4	5	6	7	Very much

25. How much were the characters looking at you?

Not at all	1	2	3	4	5	6	7	Very much

26. How much did the characters seem aware of you?

Not at all	1	2	3	4	5	6	7	Very much

27. To what extent did you feel observed by the characters?

Not at all	1	2	3	4	5	6	7	Very much

28. When you *first* saw them, did you respond to the characters essentially as if they were:

☐ People ☐ Objects (e.g. chairs) ☐ I don't know

29. Now consider your response *over the course of the whole experience.* Did you respond to the characters essentially as if they were:

☐ People ☐ Objects (e.g. chairs) ☐ I don't know

30. To what extent did you have a sense of being in the same space as the characters?

Not at all	1	2	3	4	5	6	7	Very much

31. When you *first* saw them, to what extent did you have a sense of personal contact with the characters?

Not at all	1	2	3	4	5	6	7	Very much

32. Now consider your response *over the course of the whole experience.* To what extent did you have a sense of personal contact with the characters?

Not at all	1	2	3	4	5	6	7	Very much

33. When you *first* saw the characters, was your first response to approach them or avoid them?

Approach	1	2	3	4	5	6	7	Avoid
				Neither				

34. Now consider your response *over the course of the whole experience.* Was your overall response to approach the characters or avoid them?

Approach	1	2	3	4	5	6	7	Avoid
				Neither				

35. How far did you make an effort to avoid disturbing the characters?

Not at all	1	2	3	4	5	6	7	Very much

36. How far did you feel inhibited in your task by the characters?

Not at all	1	2	3	4	5	6	7	Very inhibited

37. When you *first* saw the characters, how much did you want to interact with them?

Not at all	1	2	3	4	5	6	7	Very much

38. Now consider your response *over the course of the whole experience.* How much did you want to interact with them?

Not at all	1	2	3	4	5	6	7	Very much

39. Did you attempt to interact with them?

Not at all	1	2	3	4	5	6	7	Very much

40. Did you interact with the characters in the following ways (please check "yes" or "no" for each):

	Yes	No
Smile	☐	☐
Frown	☐	☐
Make eye contact	☐	☐
Wave	☐	☐
Touch	☐	☐
Talk	☐	☐

41. To what extent did the presence of the characters affect the way you explored the space?

Not at all	1	2	3	4	5	6	7	Very much

42. How much do you think you disturbed the characters in the room?

Not at all	1	2	3	4	5	6	7	Very much

43. When you *first* saw them, did you respond to the characters more the way you would respond to people, or the way you would respond to a computer interface?

The way I would respond to people	1	2	3	4	5	6	7	The way I would respond to a computer interface

44. Now consider your response *over the course of the whole experience.* Did you respond to the characters more the way you would respond to people, or the way you would respond to a computer interface?

The way I would respond to people	1	2	3	4	5	6	7	The way I would respond to a computer interface

Further Comments

Please write down any further comments that you wish to make about your experience. In particular, what things helped to give you a sense of 'really being' in the room, and what things acted to 'pull you out' of this? Also what tended to make the characters seem more real or take away their reality?

Reminder - all answers will be treated entirely confidentially. *Thank you once again for participating in this study, and helping with our research. Please do not discuss this with anyone for ONE MONTH. This is because the study is continuing, and you may happen to speak to someone who may be taking part.*

Semi-structured interview agenda

General
How was that?
Where were you?
What did you see?
Was there anything special/striking that stood out in the experience?
What was the first thing you noticed when you walked through the virtual doorway?

Virtual people
Can you describe them?
Can you describe their behaviour?
What was your immediate reaction?
Did your reaction change during the course of the experience?
How did they respond to you?
(Did they seem to know each other?)
Did you have a sense that you were in the company of other people?
What gave you this impression?
Did you respond to them more as: Objects or People ... People or a computer program?

Did you feel observed?
Did they seem aware of you?
Did you feel you wanted to approach or avoid them?
Were you aware of us (the researchers) when you were in the cave?

Final
If you were to tell somebody else about the experience you had here today, how would you describe it?

Appendix C

Materials for the Experiment on Eye Gaze

This appendix contains the materials used for the first experiment on eye gaze, reported in Section 5.1. These include, in order of appearance:

1. The instruction sheet given to participants at the start of each session.
2. The post-test questionnaire administered at the end of each conversation.
3. The semi-structured interview agenda.

Note that the scenarios used to prepare participants for the conversation task are included in appendix D.

Instructions

This experiment is designed to take *half an hour*. After you have signed your consent form, you will be led through to a soundproofed room and your conversation partner will be led to a separate room. The two rooms will be audio linked and both you and your conversation partner will be represented by a virtual human.

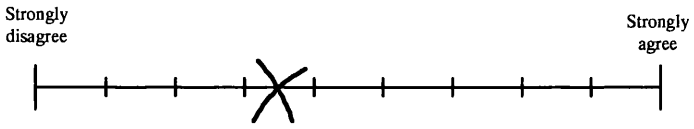
- You will be given *10 minutes* to read through a scenario. After you have finished reading and preparing, we will take the piece of paper away.
- You will then have *10 minutes* to discuss the scenario with your partner and come to an agreement.
- Finally, you will be given a brief questionnaire to fill out and you will be asked a few questions about the conversation. This is expected to take *10 minutes*.

If you have any questions, do not hesitate to ask.

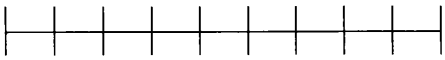
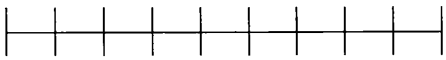
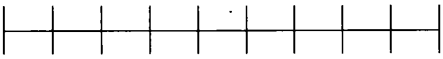
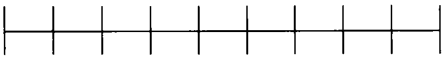
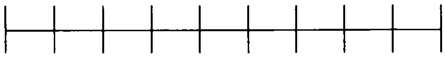
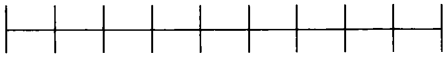
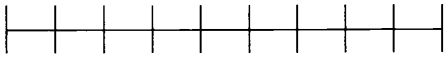
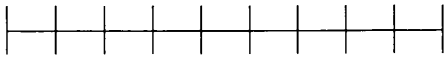
Thank you for participating!

Questionnaire

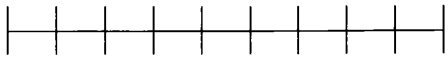
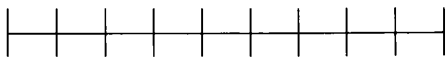
Overleaf are several statements, each followed by a bar labelled "strongly agree" on one end and "strongly disagree" at the other. For each statement, please place a cross on the line to show how you feel about the statement.



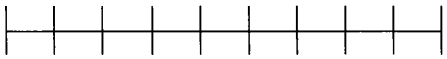
About the conversation

	Strongly Agree	Strongly Disagree
I could readily tell when my partner was concentrating on what I was saying.		
I was able to take control of the conversation when I wanted to.		
It was easy for me to contribute to the conversation.		
The conversation seemed highly interactive.		
I found it easy to keep track of the conversation.		
There were frequent inappropriate interruptions.		
This felt like a natural conversation.		
I felt completely absorbed in the conversation.		
I had a real impression of personal contact with my conversational partner.		

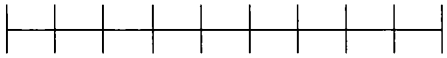
About your conversational partner

	Strongly Agree	Strongly Disagree
I was very aware of my conversational partner.		
My partner was friendly.		

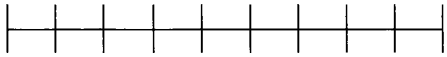
My partner did *not* take a personal interest in me.



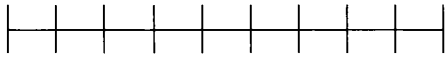
I trusted my partner.



I enjoyed talking to my partner.



I would be interested in meeting my partner face to face.



Semi-structured interview agenda

About the conversation
How did the conversation seem?
Did you enjoy it?
Did it feel comfortable?
In what ways was it like or unlike a face-to-face conversation?
To what extent did you feel absorbed in the discussion?
Was there anything in particular that you felt was missing from the conversation?

About the sense of personal contact
Did you have a real impression of personal contact with the other person?
Can you describe this?
Did you feel more like you were in the company of the other person, or more like you were alone?

About the conversation partner
Did you form an impression of your conversation partner?
What was this based on?
What effect did it have to have met them briefly beforehand?
Did you feel you trusted them?

About the visual image
Was it important that you had an image of your conversation partner in front of you?
Could you have had this conversation as easily on the telephone?
What would have been similar or different?

About the avatar
How do you think you were represented?
What did their avatar mean to you?
To what degree did you make an association between the avatar and the person you were speaking with?
Did your reaction to the avatar change during the course of the conversation?
Did you have an impression of eye contact?

Additional questions were included during the course of the experiment, based on responses from individual participants that highlighted issues of interest:

What did you focus on visually during the conversation?
Did you feel observed?
Did you have a visual image of the person you were speaking with during the conversation? What was this based on? (The scenario, the actual person, anything else?)
How could the avatar be improved?
In what way would the avatar need to be changed in order for you to consider it useful in conversation?

Appendix D

Materials for the Experiment on Eye Gaze and Photorealism

This appendix contains the materials used for the first experiment on eye gaze, reported in Section 5.2. These include, in order of appearance:

1. The instruction sheet given to participants at the start of each session.
2. The pre-questionnaire administered after participants signed the consent form.
3. The scenarios used for the roleplay negotiation task. These include male and female versions for each role (mayor and baker).
4. The post-test questionnaire administered at the end of each conversation.
5. The semi-structured interview agenda.

Note that the consent form used in this experiment was identical to the one used in the experiment on responsiveness (see Appendix B).

Instructions

Thank you for volunteering to participate in this study. This is one in a long series of studies aimed at understanding people's responses to virtual environments.

Please read through this information and feel free to ask any questions. While the experimenters will be happy to answer any general questions you may have, they have been instructed not to discuss some aspects of the study until the end.

This study will take place inside an immersive virtual reality environment. The virtual reality viewing equipment can be worn over eyeglasses if necessary. You may be asked to take off your shoes to protect the virtual reality equipment. In the virtual environment, you will meet another person and will have 10 minutes to carry out a conversation with them.

Procedures

- You will be asked to sign a Consent Form.
- You will be asked to fill out 2 preliminary questionnaires.
- You will be given 10 minutes to read through a scenario. After you have finished reading and preparing, we will take the piece of paper away.
- You will then be led through to the virtual environment.
- You will have a brief practice period to get used to moving around the virtual environment.
- Once you are comfortable, you will be asked to move into another part of the virtual environment to meet your conversation partner.
- You will then have *10 minutes* to discuss the scenario with your partner and come to an agreement.
- Afterwards you will be asked to complete a questionnaire about your experience.
- Finally, there will be a brief interview about your experience.

The whole study is expected to take *approximately 1 hour*. If you have any questions, do not hesitate to ask.

Please note

- All your questionnaire responses will remain entirely *confidential*. No identifying information about participants will be published in any form.

- Please do not discuss this study with anyone for ONE MONTH. This is because the study is continuing and you may happen to speak to someone who may be taking part.
- You are free to withdraw from the study *at any time* and *without giving reasons for withdrawing*.

A note about virtual reality equipment:

When people use virtual reality systems, some people sometimes experience some degree of nausea. If *at any time* you wish to stop taking part in the study due to this or any other reason, please say so and we will stop.

There has been some research which suggests that people using virtual reality displays might experience some disturbances in vision afterwards. No long term studies are known to us, but the studies which have been carried out do testing after about 30 minutes, and find the effect is still sometimes there.

There have been various reported side effects of using virtual reality equipment, such as "flashbacks."

With any type of video equipment there is a possibility that an epileptic episode may be generated. This, for example, has been reported for computer video games. For this reason we regret that we are unable to accept volunteers who are known to have suffered from epilepsy.

Background information

Age:

Gender:

1. Male	1
2. Female	2

My status is as follows:	*Please tick against your answer*
1. Undergraduate student	1
2. Masters student	2
3. PhD student	3
4. Research Assistant/Research Fellow	4
5. Staff member - systems/technical staff	5
6. Faculty	6
7. Administrative staff	7
8 Other (please write in)...	8

To what extent do you use a computer in your daily activities?

I use a computer...								
Not at all	1	2	3	4	5	6	7	Almost all the time

Have you experienced 'virtual reality' before?

I have experienced virtual reality...								
Never before	1	2	3	4	5	6	7	A great deal

How many times have you used an interactive virtual reality system?

I have used a virtual reality system:	*Please tick against your answer*
1. Never	1
2.	2
3.	3
4. Five times in the past year	4
5.	5
6.	6
7. On a regular basis	7

How much do you know about how virtual reality works?

I know about how virtual reality works...								
Nothing at all	1	2	3	4	5	6	7	A great deal

How much do you know about how 3D images are produced?

I know about how 3D images are produced...								
Nothing at all	1	2	3	4	5	6	7	A great deal

Please rate your level of experience with computer programming:

In terms of computer programming my level of experience is...								
Novice	1	2	3	4	5	6	7	Expert

How often do you use a video game system (at home, work, school or arcade)?

I use a video game system:	
1. Never	1
2.	2
3.	3
4. Occasionally	4
5.	5
6.	6
7. Often (several times a week)	7

How many hours per week do you spend playing video games?

I use a video game system:	
0 hours per week	1
Less than 1 hour per week	2
1 to 3 hours per week	3
3 to 5 hours per week	4
5 to 7 hours per week	5
7 to 9 hours per week	6
9 hours or more per week	7

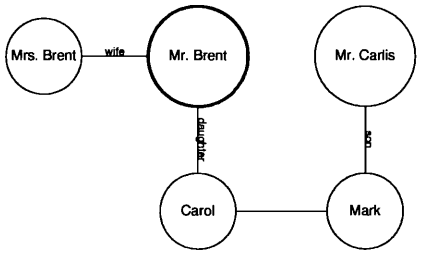
Background:

You are Mr.Brent, the owner of a small bakery. You find yourself in an extremely awkward situation and urgently need to find a solution. Your only daughter, Carol, is an aspiring actress and has been living in her own small apartment in town. In recent months she claims to have been seeing Mark Carlis, the son of the current mayor. You have never met this young man and know very little about him.

The situation is as follows: Carol claims to be pregnant by Mark. She explains that she cares about Mark deeply and wants to keep the child. However, Mark has not offered to marry her or to provide financial support in any form. He was shocked when Carol gave him the news a few days earlier and has not been in touch with her since. You realise the pregnancy would be a terrible scandal for your family. You also do not want Carol to face a future as a single mother. At the moment she is not earning well, and you and your wife certainly can't do much to help as you yourselves are in hard times financially.

Your wife has been extremely upset about the situation. She has written a letter to the newspaper about the story, which she is threatening to send. A scandalous story involving the son of the current mayor is something the papers would be more than happy to run, but you can see no advantage in letting this happen. You know you will receive no financial reward from the paper for giving them the story. It will only hurt Carol, and will almost certainly alienate the mayor and his son. They may, after all, be willing to help. You have tried to persuade your wife not to send the letter, but she has told you that unless you can come up with a better solution in the next half hour she will mail it to the editor-in-chief.

You have phoned the mayor and have quickly outlined the situation and have offered to meet him immediately to talk this through. He has agreed. You realise that it is within everybody's interests to prevent the letter from reaching the press, as the mayor must wish to avoid a scandal. By the time you arrive, you have only ten minutes left in which to come to a solution.



The task:

You and your conversation partner have *10 minutes* to come to an agreement. Your conversation partner will be representing Mr. Carlis, the mayor. If by the end of the ten minutes

you have not come up with an alternative, your wife will mail the letter to the press.

Remember that you need to be as diplomatic as possible because you need to come to a mutually acceptable decision and you don't have much time.

Below are three possible alternatives you may choose to discuss

There are THREE ALTERNATIVES:

1. *The mayor might try to argue that you have no case*-this would be the worst-case scenario, as your wife would then send the letter and you would then have to way of helping Carol. You also absolutely want to avoid a scandal that would bring further unhappiness to your family.

2. *Come to some form of financial settlement*-perhaps Mark could offer to help support the baby financially. However, you don't know how the mayor would react to a request for money and you certainly don't want to think you are trying to blackmail him. More importantly, Carol would remain without a husband and would be unhappy.

3. *Let them marry*-this would certainly be the best solution. Carol wants to keep the child and be with Mark. You want to see her happy, and your wife would be relieved to see her settled.

Background:

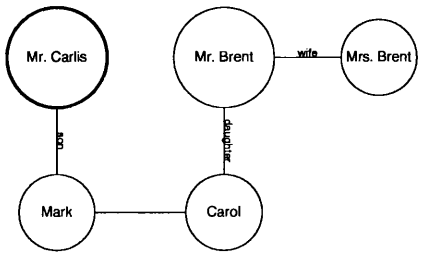
You are Mr. Carlis, the mayor of a small town. You receive a telephone call from a Mr.Brent, who claims he needs to see you urgently on personal business.

He explains that your respective children have involved both of you in a very awkward situation. His wife, Mrs. Brent, is so upset about it that she has written a letter to the newspaper outlining the whole story. Though he has tried to dissuade her from doing this, arguing that it will do nobody any good, she has threatened to send the letter within the next half hour unless he and you can come up with a better solution.

The situation is as follows: Mr. Brent and his wife own a small bakery in town. They have an only daughter, Carol, who is an aspiring actress. Carol has been seeing your son Mark for the last few months, and she has just told her parents that she is carrying Mark's child. She is apparently deeply upset because she has not heard from Mark since telling him the news two days ago.

You were not aware that Mark was in a relationship with this young woman, and are caught by surprise. No matter what the truth of the situation is, you must prevent this letter from reaching the press because it would almost certainly be blown out of all proportion. Mark has a reputation to uphold in the legal profession. You want him to become a high court judge, and you certainly don't want to jeopardise his chances because of this potential scandal. Also, your own mayoral election is coming up and you can't afford any negative publicity.

You agree with Mr. Brent that the two of you meet immediately. In the few intervening minutes, you contact your son. He confirms that he has had seen Carol several times, but explains that they are in no way in an official relationship; he certainly has no intention of marrying her. He also explains that he has good reason to believe he is not the only man in her life. Your son has never been in this kind of trouble before, but you know he lives alone now and leads his own life. You understand that it's within everybody's interests to prevent the letter from reaching the press. It will cause a scandal for both families, and in any case Mrs. Brent will receive no financial reward for telling the papers. By the time you and Mr. Brent meet, you have only ten minutes left in which to come to a solution.



The task:

You and your conversation partner have 10 minutes to come to an agreement. Your conversation partner will be representing Mr. Brent. If by the end of the ten minutes you have not come up with a solution, his wife will mail the letter to the press.

Remember that you need to be as diplomatic as possible because you need to come to a mutually acceptable decision and you don't have much time.

Below are three possible alternatives you may choose to discuss

There are THREE ALTERNATIVES:

1. *Let them marry*-this is the least acceptable solution for you, as your son Mark has already explained he does not want this.

2. *Come to some form of financial settlement*-this is also not an ideal solution. In any case, Mr. Brent has made no mention of money, and he might be deeply offended if you were to suggest this.

3. *Argue that Mr. Brent doesn't have a case*-this is the preferable solution, although you wish to avoid a deadlock. You absolutely cannot afford to let this letter go to the press.

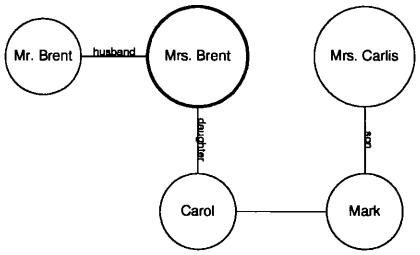
Background:

You are Mrs. Brent, the owner of a small bakery. You find yourself in an extremely awkward situation and urgently need to find a solution. Your only daughter, Carol, is an aspiring actress and has been living in her own small apartment in town. In recent months she claims to have been seeing Mark Carlis, the son of the current mayor. You have never met this young man and know very little about him.

The situation is as follows: Carol claims to be pregnant by Mark. She explains that she cares about Mark deeply and wants to keep the child. However, Mark has not offered to marry her or to provide financial support in any form. He was shocked when Carol gave him the news a few days earlier and has not been in touch with her since. You realise the pregnancy would be a terrible scandal for your family. You also do not want Carol to face a future as a single mother. At the moment she is not earning well, and you and your husband certainly can't do much to help as you yourselves are in hard times financially.

Your husband has been extremely upset about the situation. He has written a letter to the newspaper about the story, which he is threatening to send. A scandalous story involving the son of the current mayor is something the papers would be more than happy to run, but you can see no advantage in letting this happen. You know you will receive no financial reward from the paper for giving them the story. It will only hurt Carol, and will almost certainly alienate the mayor and his son. They may, after all, be willing to help. You have tried to persuade your husband not to send the letter, but he has told you that unless you can come up with a better solution in the next half hour he will mail it to the editor-in-chief.

You have phoned the mayor and have quickly outlined the situation and have offered to meet her immediately to talk this through. She has agreed. You realise that it is within everybody's interests to prevent the letter from reaching the press, as the mayor must wish to avoid a scandal. By the time you arrive, you have only ten minutes left in which to come to a solution.



The task:

You and your conversation partner have *10 minutes* to come to an agreement. Your conversation partner will be representing Mrs. Carlis, the mayor. If by the end of the ten minutes

you have not come up with an alternative, your husband will mail the letter to the press.

Remember that you need to be as diplomatic as possible because you need to come to a mutually acceptable decision and you don't have much time.

Below are three possible alternatives you may choose to discuss

There are THREE ALTERNATIVES:

1. *The mayor might try to argue that you have no case*-this would be the worst-case scenario, as your husband would then send the letter and you would then have no way of helping Carol. You also absolutely want to avoid a scandal that would bring further unhappiness to your family.

2. *Come to some form of financial settlement*-perhaps Mark could offer to help support the baby financially. However, you don't know how the mayor would react to a request for money and you certainly don't want to think you are trying to blackmail her. More importantly, Carol would remain without a husband and would be unhappy.

3. *Let them marry*-this would certainly be the best solution. Carol wants to keep the child and be with Mark. You want to see her happy, and your husband would be relieved to see her settled.

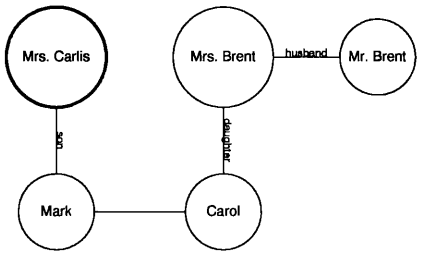
Background:

You are Mrs. Carlis, the mayor of a small town. You receive a telephone call from a Mrs.Brent, who claims she needs to see you urgently on personal business. She explains that your respective children have involved both of you in a very awkward situation. Her husband, Mr. Brent, is so upset about it that he has written a letter to the newspaper outlining the whole story. Though she has tried to dissuade him from doing this, arguing that it will do nobody any good, he has threatened to send the letter within the next half hour unless she and you can come up with a better solution.

The situation is as follows: Mrs. Brent and her husband own a small bakery in town. They have an only daughter, Carol, who is an aspiring actress. Carol has been seeing your son Mark for the last few months, and she has just told her parents that she is carrying Mark's child. She is apparently deeply upset because she has not heard from Mark since telling him the news two days ago.

You were not aware that Mark was in a relationship with this young woman, and are caught by surprise. No matter what the truth of the situation is, you must prevent this letter from reaching the press because it would almost certainly be blown out of all proportion. Mark has a reputation to uphold in the legal profession. You want him to become a high court judge, and you certainly don't want to jeopardise his chances because of this potential scandal. Also, your own mayoral election is coming up and you can't afford any negative publicity.

You agree with Mrs. Brent that the two of you should meet immediately. In the few intervening minutes, you contact your son. He confirms that he has had seen Carol several times, but explains that they are in no way in an official relationship; he certainly has no intention of marrying her. He also explains that he has good reason to believe he is not the only man in her life. Your son has never been in this kind of trouble before, but you know he lives alone now and leads his own life. You understand that it's within everybody's interests to prevent the letter from reaching the press. It will cause a scandal for both families, and in any case Mr. Brent will receive no financial reward for telling the papers. By the time you and Mrs. Brent meet, you have only ten minutes left in which to come to a solution.



The task:

You and your conversation partner have *10 minutes* to come to an agreement. Your conversation partner will be representing Mrs. Brent. If by the end of the ten minutes you have not come up with a solution, her husband will mail the letter to the press.

Remember that you need to be as diplomatic as possible because you need to come to a mutually acceptable decision and you don't have much time.

Below are three possible alternatives you may choose to discuss

There are THREE ALTERNATIVES:

1. *Let them marry*-this is the least acceptable solution for you, as your son Mark has already explained he does not want this.

2. *Come to some form of financial settlement*-this is also not an ideal solution. In any case, Mrs. Brent has made no mention of money, and she might be deeply offended if you were to suggest this.

3. *Argue that Mrs. Brent doesn't have a case*-this is the preferable solution, although you wish to avoid a deadlock. You absolutely cannot afford to let this letter go to the press.

Virtual Reality Study Questionnaire

Below are several statements, each followed by a scale labelled "strongly agree" on one end and "strongly disagree" at the other. For each statement, please select a number from 1 to 7 to show how you feel about the statement.

	Strongly Disagree						Strongly Agree
I could readily tell when my partner was concentrating on what I was saying	1	2	3	4	5	6	7
I was able to take control of the conversation when I wanted to	1	2	3	4	5	6	7
It was easy for me to contribute to the conversation	1	2	3	4	5	6	7
The conversation seemed highly interactive	1	2	3	4	5	6	7
I found it easy to keep track of the conversation	1	2	3	4	5	6	7
There were frequent inappropriate interruptions	1	2	3	4	5	6	7
This felt like a natural conversation	1	2	3	4	5	6	7
I felt completely absorbed in the conversation	1	2	3	4	5	6	7
I had a real impression of personal contact with my conversational partner	1	2	3	4	5	6	7
I was very aware of my conversational partner	1	2	3	4	5	6	7
My partner was friendly	1	2	3	4	5	6	7
My partner did *not* take a personal interest in me	1	2	3	4	5	6	7
I trusted my partner	1	2	3	4	5	6	7
I enjoyed talking to my partner	1	2	3	4	5	6	7
I would be interested in meeting my partner face to face	1	2	3	4	5	6	7
This felt like a phone conversation	1	2	3	4	5	6	7
I felt alone	1	2	3	4	5	6	7
I did *not* feel my partner and I were together	1	2	3	4	5	6	7
I had a sense of being in the company of my conversation partner	1	2	3	4	5	6	7
I behaved as if there was nobody watching me	1	2	3	4	5	6	7
The interaction did not seem very personal to me	1	2	3	4	5	6	7
It was as if I had a person in front of me	1	2	3	4	5	6	7
I felt my partner and I were in a shared space	1	2	3	4	5	6	7
Most of my attention was focused on my partner's voice	1	2	3	4	5	6	7
My partner and I frequently made eye contact	1	2	3	4	5	6	7
The way my partner looked at me appeared natural	1	2	3	4	5	6	7
I had a sense of being with the other person	1	2	3	4	5	6	7
It did *not* feel as though my partner and I were in the same room	1	2	3	4	5	6	7
It seemed as if my partner was observing me	1	2	3	4	5	6	7
I behaved as if there was nobody in the virtual room with me	1	2	3	4	5	6	7

Did you attempt to interact with your partner in any of the following ways?

I tried to interact with my partner in the following ways:	*Please tick against your answer(s)*
1. Smile	1
2. Shake hands	2
3. Wave	3

In virtual environments, we call virtual humans *avatars*. Below are several statements about the avatar in the virtual environment you just experienced. For each statement, please select a number from 1 to 7 to show how you feel about the statement.

	Strongly Disagree						Strongly Agree
The avatar had *no* impact on the conversation	1	2	3	4	5	6	7
I paid less attention to the avatar as the conversation went on	1	2	3	4	5	6	7
The avatar's appearance was realistic	1	2	3	4	5	6	7
The avatar's behaviour was realistic	1	2	3	4	5	6	7
The avatar did *not* always seem to respond appropriately to me	1	2	3	4	5	6	7
I believe the avatar must have resembled my conversation partner	1	2	3	4	5	6	7
The avatar was *not* expressive	1	2	3	4	5	6	7

Please rate the extent to which the avatar *helped* you to:

	Not at all						Very much
Understand how your partner was feeling	1	2	3	4	5	6	7
Understand your partner's reactions to your conversation	1	2	3	4	5	6	7
Form a clear impression of your partner	1	2	3	4	5	6	7
Build a sense of personal contact with your partner	1	2	3	4	5	6	7
Get to know your partner	1	2	3	4	5	6	7
Have a natural conversation	1	2	3	4	5	6	7
Trust your partner	1	2	3	4	5	6	7
Come to an agreement with your partner	1	2	3	4	5	6	7

When you *first* saw the avatar, to what extent did it become the person you were meeting with?

Not at all	1	2	3	4	5	6	7	Very much so

Now consider your response *over the course of the whole experience*. To what extent did it become the person you were meeting with?

Not at all	1	2	3	4	5	6	7	Very much so

In the beginning of the conversation, to what extent did you pay attention to the avatar?

Not at all	1	2	3	4	5	6	7	Very much so

Now consider your response over the course of the whole experience. To what extent did you pay attention to the avatar?

Not at all	1	2	3	4	5	6	7	Very much so

To what degree do you think the avatar's actions were controlled by your partner?

Not at all	1	2	3	4	5	6	7	Very much so

To what degree do you think the avatar's actions reflected what your partner was actually doing?

Not at all	1	2	3	4	5	6	7	Very much so

Did the avatar seem more like a person, or more like a computer?

Like a person	1	2	3	4	5	6	7	Like a computer

Please note that in the final part of this questionnaire "laboratory" refers to the physical space in which the study is taking place, and "room" refers to the virtual room you were in.

Please rate the extent to which you were aware of background sounds in the laboratory in which this experience was actually taking place. Rate this on the following scale from 1 to 7 (where for example 1 means that you were not at all aware of the background sounds):

During the experience I was thinking that I was really in laboratory...								
Not at all	1	2	3	4	5	6	7	Very much so

How dizzy, sick or nauseous did you feel resulting from the experience, if at all? Please answer on the following 1 to 7 scale.

I felt sick or dizzy or nauseous during or as a result of the experience...								
Not at all	1	2	3	4	5	6	7	Very much so

Please rate your *sense of being in the* room, on the following scale from 1 to 7, where 7 represents your *normal experience of being in a place.*

I had a sense of being there in the room...								
Not at all	1	2	3	4	5	6	7	Very much so

To what extent were there times during the experience when the room was the reality for you and you almost forgot about the real world of the laboratory where the experience was really taking place?

There were times during the experience when the room was the reality for me...								
At no time	1	2	3	4	5	6	7	Almost all of the time

When you think back about your experience, do you think of the room more as *images that you saw*, or more as *somewhere that you visited*?

The room seems to me to be more like...								
Images that I saw	1	2	3	4	5	6	7	Somewhere I visited

During the course of the experience, which was strongest on the whole, your sense of being in the room, or of being in the real world of the laboratory?

I had a stronger sense of...								
Being in the lab	1	2	3	4	5	6	7	Being in the room

Overall, how well do you think that you achieved your task?

I achieved my task...								
Not at all	1	2	3	4	5	6	7	Fully

During the time of the experience, did you often think to yourself that you were just standing in a laboratory or did the room overwhelm you?

During the experience I was thinking that I was really in laboratory...								
Most of the time	1	2	3	4	5	6	7	Rarely

During the course of the experience, how much were you aware of the experimenters?

Not at all	1	2	3	4	5	6	7	Very much

Further Comments

Please write down any further comments that you wish to make about your experience. **Reminder - all answers will be treated entirely confidentially.** *Thank you once again for participating in this study, and helping with our research. Please do not discuss this with anyone for ONE MONTH. This is because the study is continuing, and you may happen to speak to someone who may be taking part.*

Semi-structured interview agenda

General

How was that?

Where were you?

What did you see?

Was there anything special or striking that stood out in the experience?

What was the first thing you noticed when you walked through the doorway into the virtual room?

If you were to tell somebody else about the experience you had here today, how would you describe it?

What did you focus on during the conversation?

What did you focus on visually during the conversation?

Related to copresence

Did you have an impression of eye contact?

Did you have a sense of personal contact with your conversation partner?

Can you describe this?

Did you have a sense of being in the company of the person/alone?

Did you have a sense of being in the same space, or different spaces?

About the conversation partner

Did you form an impression of your partner?

What was this based on?

Did you feel you trusted him/her?

What was this based on?

Did you feel observed?

About the conversation

How was the conversation?

Did you enjoy it?

Did it feel comfortable?

In what ways was it like or unlike a face-to-face conversation?

Did you feel absorbed in the discussion?

Was there anything in particular you felt was missing from the conversation?

Could you have had this conversation as easily on the telephone?

What would have been similar/different?

About the avatar

Was it important to have the avatar there?

Did you respond to it more as a person or a computer?

Did your reaction to the avatar change during the conversation?

Final

What do you think your partner looks like in real life?

Appendix E

Further Analysis for the Experiment on Gaze and Photorealism

This appendix contains additional results from the analysis of questionnaire responses from the experiment on eye gaze and photorealism (reported in Section 5.2). Four response variables are considered. The questionnaire items making up each variable are detailed in Section E.1, and the findings are presented in Section E.2.

E.1 Additional response variables

1. **As person** (n=2): the extent to which the avatar was perceived as real and like a human.
   (a) It was as if I had a person in front of me.
   (b) Did the avatar seem more like a person, or more like a computer?
2. **Association** (n=5): The degree to which the avatar was associated with the conversation partner:
   (a) To what degree do you think the avatar's actions were controlled by your partner?
   (b) To what degree do you think the avatar's actions reflected what your partner was actually doing?
   (c) I believe the avatar must have resembled my conversation partner
   (d) When you *first* saw the avatar, to what extent did it become the person you were meeting with?
   (e) Now consider your response *over the course of the whole experience.* To what extent did it become the person you were meeting with?
3. **Attention** (n=3): The degree to which participants paid attention to the avatar in the course of the experience:

(a) *In the beginning of the conversation,* to what extent did you pay attention to the avatar?

(b) Now consider your response *over the course of the whole experience.* To what extent did you pay attention to the avatar?

(c) I paid less attention to the avatar as the conversation went on

4. **Avatar contribution** (n=8): The degree to which the avatar helped to:

(a) Understand how your partner was feeling.

(b) Understand your partner's reactions to your conversation.

(c) Form a clear impression of your partner.

(d) Build a sense of personal contact with your partner.

(e) Get to know your partner.

(f) Have a natural conversation.

(g) Trust your partner.

(h) Come to an agreement with your partner.

E.2 Results

Table E.1: Logistic regression for the additional count response variables

The tabulated χ^2 5% value is 3.841 on 1 d.f., and all d.f.s are 1.
The deviance column shows the increase in deviance that would result if the corresponding variable were deleted from the model.

	Deviance χ^2			
Fitted Variable	**As person**	**Association**	**Attention**	**Avatar contribution**
Type gaze (inferred)	-	-	8.29	-
Age	10.34 (+)	5.75 (+)	12.73 (+)	10.12 (+)
Gender (female)	-	-	-	15.59
SAD	-	5.51 (-)	14.55 (-)	-
VR experience	-	-	4.36 (-)	-
VR times	-	-	-	11.01 (-)
Programming	-	-	-	10.51 (+)
Game times	-	-	-	11.37 (+)
Overall deviance	59.74	74.22	53.79	105.3
Overall d.f.	46	44	42	42

Table E.1 shows the results of the logistic regression on each of the additional response variables. *Age* significantly affected the degree to which the avatar was seen *as a person,* with

older participants more likely to see it as a person than as a computer interface. Both participants' age and SAD score had a significant effect on *association*. Older participants were more likely to associate the avatar in front of them with their conversation partner, while participants with higher social anxiety scores were less likely to make a close association between partner and avatar.

The type of gaze affected the amount of attention participants paid to the avatar, with the inferred-gaze model resulting in a lower count. Again, age positively affected the response, with older participants paying more attention to the avatar. However, higher SAD scores and greater experience with VR systems both significantly reduced the amount of attention paid to the avatar.

The final variable concerned the degree to which the avatar was perceived as making a positive contribution to the interaction in a number of ways. *Age*, again, had a significant positive effect. Females were also significantly more likely to see the avatar as contributing to the interaction. Both level of *programming* experience and number of times participants played computer games had a significant positive effect on this response. However, participants with a greater number of exposures to virtual reality systems were significantly less likely to see the avatar as contributing to the interaction.

Appendix F

Grounded Theory Analysis

This appendix contains documents relevant to the grounded theory analysis reported in Chapter 6. These include, in order of appearance:

1. A sample interview transcript from a participant in the *inferred-gaze* avatar condition.
2. The list of codes generated in the analysis.
3. A summary of the process of visual abstraction used for the final stages of selective coding.

F.1 Sample Interview Transcript

C3G1B **(Male playing Baker role)**

Experimenter	how did the conversation feel
Participant B	the actual voice part of it or the picture as well or the whole thing
Experimenter	the whole thing
Participant B	fine yeah fairly easy I could hear him properly I could see what I thought was a fair representation of him nodding and eye contact was not quite as I would have expected it but yeah it was alright
Experimenter	right ok were you comfortable talking to someone using this
Participant B	yes fairly yes it wasn't a problem yeah I'd do it again
Experimenter	ok so how did you feel using this was like or unlike a real life face to face conversation
Participant B	um (pause) it was I could probably have had the same conversation without seeing the actual face in front of me I suppose in some ways that's a negative point of view to it um I think I got my point of view across verbally rather than actually focusing on what he was doing or nodding or shaking the head I suppose the agreements were there with him shaking the head or nodding and shaking his head and he did sort of counteract that with the voice as well and I think probably we were using our voices more than actually using the picture
Experimenter	right so how would you compare this to a phone conversation if you'd done this on the phone
Participant B	yeah it would probably have taken us the same length of time on the phone so I suppose in some ways this was probably more friendly in the sense that if I had actually acknowledged that the so called human that I'd seen in front of me was actually a person I suppose I was looking at him and thinking yeah ok there's a representation of the person but actually I understand the voice so the voice takes the sort of main role cause that's what you understand (pause) was it useful to have him I suppose in some ways yes it was useful because you could see the nodding in agreement with you and disagreeing or I don't think he actually did disagree with me in our conversation but I suppose it was useful from that point of view

Experimenter alright you mentioned the nodding did you get any other kind of feedback from it or did you notice any other meaningful any other behaviours

Participant B no I didn't notice any other behaviours the eyes did strange things and I don't quite know why I just saw whites rather than the eyeball sometimes um you could see him twist his head when he was looking in other directions but apart from that no there wasn't any other movement as far as I know

Experimenter ok was there anything in particular that you felt was missing from this conversation that would have been there in a face to face conversation

Participant B I didn't feel so pressurised in this particular conversation because as I might have done if it was a face to face in the sense that I knew that person couldn't come and sort of slap me around the face in this particular conversation if they'd felt so aggrieved because of the scenario we were in that he wanted to slap me around the face so from that point of view I didn't feel under any pressure I suppose that's the only difference really and that was fine

Experimenter right ok was it important for you to have this image in front of you

Participant B (pause) no sorry to say as I said I think I could have had the same conversation along the same sort of lines taking about the same length of time to understand and divulge the information verbally rather than actually viewing as well as verbally

Experimenter verbally do you mean

Participant B just via a normal ok just via a phone call

Experimenter ok so you feel with this you got no kind of

Participant B the only feedback I had was the nodding in agreement um but he also then said yes we can make this work and he agreed with me verbally that he could make this work for us so the nodding in agreement was the only thing I think that helped so if you want some positive feedback to it then yes there was that sort of interaction there between the two of us but I still think I could have had the same conversation over the phone

Experimenter right did you find yourself trying to make eye contact at all with the avatar

Participant B yes yes I did and I didn't feel that I was getting any responses back the normal sort of eye contact that I'd get like looking at yourself now I didn't feel that I was having that acknowledgement back as it would be in a normal conversation

Experimenter ok so did you have any impression of personal contact with the other person

Participant B yes if you take the nodding then yes that was the personal feedback and contact that I had with that person yeah but there was nothing else

Experimenter ok did you feel you were in their company or did you feel you were alone and isolated

Participant B (garbled) the avatar was male and the voice was male the actual nodding with agreement it was as though it was a normal conversation the only thing as I say that made it unrealistic was the lack of eye contact that we probably had but apart from that then I suppose it was fairly realistic to the conversation that we were having

Experimenter ok did you form an impression of your conversation partner

Participant B yeah he was willing to listen to me we came to a conclusion which is what we were asked to do and I think if I hadn't built an impression of him or come across the impression of him that I'd built in my mind then we probably wouldn't have come to a conclusion so yeah I think I did have an impression of him

Experimenter right on what did you base this impression

Participant B probably as the avatar (garbled) to the picture that I got in front of me but apart from that then yeah

Experimenter did you see the other guy at all beforehand

Participant B yeah very briefly

Experimenter yeah did this have any effect on how you saw the avatar

Participant B no I don't think it did his voice the impression that I think I probably built of that person very quickly probably within two seconds of seeing him the voice backed up my impression and I thought yes this is the person that I saw five minutes ago ten minutes ago um but then you then put the voice to the avatar and it didn't really bother me to be honest

Experimenter ok do you think if you hadn't seen him beforehand do you think this would have affected how you saw the avatar at all

Participant B (pause) no because I think you'd still build an impression from the voice tones of the human being rather than in the split second that I saw him so I think the voice gave me the impression of what that person actually looks like the avatar didn't give me the physical impression that I'd built in my mind of that person yeah

Experimenter ok did you feel you trusted him

Participant B yeah I felt everything that he said I trusted

Experimenter ok

Participant B you're now going to tell me the mayor is bankrupt (laughs)

Experimenter he could have been

Participant B (laughs)

Experimenter did you trust him any differently than say if you had met him face to face

Participant B no I don't think so you build up ok if you put the voice that I heard to the avatar then that was the person that I was building that relationship with and trusting and we came to the conclusion that we came to I think in further conversations if he had then not done what he said he was going to do then I would start to distrust him but in the particular conversation that we were having I would say that I trusted what I heard and saw

Experimenter ok how do you think that you were represented to the other person

Participant B (laughs) if it was the same person I saw in front of me I would have hoped that he saw more nodding um cause I'm probably a fairly movable character as I sort of speak I tend to sort of create more moves so I hope that he saw a sort of more acknowledging character in front of him if he saw a blonde female then he's going to be quite confused by the voice and everything else so I would hope that what I saw he saw from character appearance

Experimenter right so did you base this just on what you saw

Participant B yeah I think so

Experimenter ok what did the avatar mean to you

Participant B the mayor was in his pyjamas (laughs) what did that avatar mean to me (pause) not a lot I suppose it represented the person that I was talking to it didn't really give me anything else didn't give me any feeling of wealth importance anything like that I suppose it actually made the mayor as a character look quite feeble and sort of quite weeny in body proportions that sort of thing but I suppose that's what it gave to me the impression that it gave to me anyway

Experimenter ok was there any way that you think it could have been improved

Participant B yeah I think it could have been dressed as I would imagine a mayor to be dressed with the sort of chain around him maybe a smart suit or something like that maybe an office in the background rather than a blue screen it would have given a more realistic impression of where I envisaged he would be because I think it said in the blurb that he was in his office and I was quite close to the office at the time so yeah I'd imagine a sort of state room or something in the background which would have made it more realistic

Experimenter ok what about the behaviours and actions of the avatar

Participant B ok I didn't see any arm movement I suppose you do see some arm movement in a normal realistic conversation the mouth didn't move very much as he spoke and didn't seem to change shape sort of lengthwise or widthwise it just opened and closed um did I see blinking I probably did I might have seen some blinking I suppose that's quite realistic and the eyes didn't really give me an awful lot of eye contact the movement from side from side if he was moving his face from side to side at the time of doing it then that was quite realistic

Experimenter ok do you think if those things you mentioned had been incorporated into avatars do you think that would give you more feedback and enriched your conversation

Participant B yeah definitely and I'd have probably where I said earlier that I didn't feel that I was under any pressure because I was speaking to an avatar rather than a human being I'd think that I'd have probably felt yes I'm speaking to a mayor now rather than someone in his pyjamas and I'd have felt I've got to actually put my point of view across in a more diplomatic way although it told me to put it forward in a more diplomatic way I'd have been probably been a bit more aware of what I was saying before having said it rather than just sitting here quite relaxed saying what do you think you're doing your son's gone and got my daughter pregnant I'd have sort of sat upright a bit more and yeah gotten my point of view across probably a bit better

Experimenter ok so any other associations that you made between the avatar and the person rather than say dissociations due to a lack of mayor-like appearance

Participant B (pause) no I don't think so no as I said the only thing was the actual background itself rather than being in an office any associations I made no I don't think there were

Experimenter ok how did you find yourself treating the avatar

Participant B I was probably trying well I was aware of his movements and that sometimes when I was speaking to him he was not looking directly at me he was looking in a slightly different direction which I suppose made me think concentrate on what I'm saying mr mayor rather than obviously being distracted by something else but it may have been that he wasn't actually looking somewhere else he was meant to be looking at me but I don't know so yeah I found myself looking at him yeah I suppose I was concentrating on his movements seeing whether he did acknowledge me as I sort of asked questions or put my point of view across which sometimes he did

Experimenter ok did you find that your behaviour towards the avatar changed throughout the conversation

Participant B no don't think so I think I knew that I was going to be talking to well I didn't know I was going to be talking to an avatar but once it appeared on the screen I knew I had ten minutes to get down to the task in hand so I spoke to the avatar as he was the person that was representing the mayor yeah that didn't change throughout I don't think

Experimenter ok what do you think kept your attention

Participant B a mixture of the voice the conversation and the few movements that the avatar gave the occasional nod in the right place but I'd have had my attention kept more if the eyes gave me a more sort of eye to eye contact and he'd acknowledged in bigger nods and he'd actually spoken as I would normally have expected a sort of normal speaking person's mouth to move so yeah that's the basics around that one

Experimenter ok I think I've dealt with all my points here any other comments or questions that you want to make or ask

Participant B did he see the same picture as I did

Experimenter yes he did

Participant B I don't have any other comments it was good it was quite good I think that the main thing for me was I'd have liked to have seen a more realistic background in the back of the avatar but that's only a personal impression because what someone deems as a background to where a mayor should be sitting in his office could be different for someone else so that's quite a difficult thing to talk about and decide

Experimenter yeah do you think it would have been different maybe if there was just a blank say just a white background

Participant B possibly I would have thought he was in a photo booth or something like that (pause) yeah I don't know how that would have changed it I suppose the blue was quite a subtle colour rather than anything else did he have the same colour background as well or

Experimenter yeah

Participant B yeah for me personally I'd have liked to have seen a room or something physical behind him maybe a plant or something like that rather than just a plain colour that would have been more pleasing to me to my eyes anyway

Experimenter ok

F.2 Codes Generated in the Analysis

- abruptness
- absent
- absorption
- absorption %high
- absorption %low
- abstract
- action
- activity %limited
- actual
- added value
- added value %limited
- added value %none
- added value %reduced
- addressing
- age
- alive
- alone physically %high
- alone physically %low
- alone rationally
- ambiguity %reduced
- ambiguous
- amount
- amplitude %high
- amplitude %low
- animation
- animation %limited
- animation %quantity
- animation %selective
- anonymity
- anonymity %partial
- appearance
- appearance %unconvincing
- appearance control
- appearance mirroring
- appropriateness %low
- appropriateness
- appropriateness %periodical
- artificial
- associating partner with avatar
- association
- association %low
- assumption
- attention
- attention %divided
- attention %increased
- attention %receiving
- attention %seeking
- attire %inappropriate
- attitude
- attitude %helpful
- attitude %negative
- audio
- audio feedback
- audio quality %high
- audio tracking
- audiovisual %consistency
- audiovisual %discrepancy
- avatar
- avatar %generic
- awareness
- awareness %high
- awareness %low
- awareness mediation %high
- background
- background %generic

- background %personal
- background %technical
- balancing
- banter
- beginning process
- behaviour %falling in
- behavioural modification
- behavioural realism
- behaviours
- belief
- believability %low
- body language
- body movement
- boredom
- breaking ice
- brevity
- building
- business
- character
- clarification needed
- clarity %high
- clarity %low
- clothes
- cognitive load
- co-location
- co-location %low
- comfort %high
- comfort %low
- communication
- communication success
- communicative value
- company
- company %low
- comparison %different
- comparison %similar
- compatibility
- compromise %none
- computer generated
- concentration %loss
- concentration %low
- confrontational %high
- confrontational %low
- confusion
- congruence %none
- conscious
- consistency
- context
- continuity
- contradiction
- contrived
- conversation flow
- conversation
- conversation %diplomatic
- conversation %forced
- conversation %mediated
- conversation %unchanged
- conversation during
- conversation impression
- conversation partner
- conversation quality %good
- conversational approach
- conversational quality
- copresence
- copresence %low
- cues
- curiosity
- decoding

- delay
- deliberate deception
- demand characteristics
- detachment
- difficult
- difficult %low
- diplomatic %high
- direct communication
- disconcerting
- disjoint
- distance
- distraction
- distraction %low
- effective coincidence
- effort expended
- effort expended %none
- email
- emotion masking
- emotional
- emotional interaction
- emotions
- empty
- enjoyment %high
- equilibrium
- equivocal %high
- equivocal %low
- erratic
- expectation
- expectations %failed
- expectations %met
- expectations %unmet
- experienced user
- express interest
- express interest %none
- expression %interest
- expression %neutral
- expressions
- expressions %none
- expressions %vague
- eye contact
- eye contact %frequent
- eye contact %none
- eye contact %periodical
- eye movement
- eye movement %away
- eye movement %upwards
- eye movement %wandering
- eyebrow
- eyebrows
- eyes
- face
- face to face
- facial expression
- faithfulness %uncertain
- familiarity
- familiarity %low
- features
- feedback
- feedback %giving
- feedback %limited
- feedback %none
- feedback %uncertain
- feelings %unknown
- focus
- focus %low
- focus %mental
- foreground
- formalities
- frequency
- gaze

- gaze %interested
- gaze %sending
- gaze %staring
- gaze %while listening
- gaze %while speaking
- gaze avoidance
- gaze behaviour
- gaze tracking
- gesture
- guarantee
- habituation
- haircut
- hand animation
- hand gestures
- hear self %low
- honesty
- human presence %low
- humanlike
- ignore
- impact %negative
- imaginary space
- impact
- impact %amusing
- impact %high
- impact %low
- impact %negative
- impact %neutral
- impact %none
- impact %positive
- impact %unknown
- impedes
- importance %high
- importance %low
- impression formation
- impression formation %negative
- impression formation %none
- impressionistic
- improvement
- incidental
- independent movement
- individuality
- information
- information %insufficient
- inhibition
- inhibits
- insufficient
- intention
- interaction
- interaction %diminished
- interaction %natural
- interest
- interest %low
- interpersonal distance %excessive
- interpretation
- intimacy
- intimacy %low
- intimacy
- intrusion
- investment %low
- isolation
- intrude
- knowing partner
- knowledge
- lacking
- leaning in
- lifelike %low
- limitation
- lip-synch
- lip-synch %shapes

- listening
- logical distinction
- lonely %low
- meaning %none
- meaning %unknown
- mediation
- meeting beforehand
- mental image
- mirroring
- mirroring %certain
- mirroring %none
- mirroring %uncertain
- mirroring monitoring
- misleading
- mood %hesitant
- mood %hostile
- mood %playful
- mood %serious
- mouth
- mouth shapes
- movement
- movement %constant
- movement %dynamic
- movement %excessive
- movement %limited
- movement %natural
- movement %static
- movement %subtle
- movement %swaying
- movement simulation
- mutual gaze
- mutual gaze %limited
- mutual gaze %none
- natural
- natural %low
- need satisfaction
- negotiation strategy
- network reliability
- nonverbal
- novelty
- obligation
- observation
- observation %neutral
- observed
- observed %low
- observed %uncertain
- obstacle
- off-putting
- one to one
- openness
- outcome
- overrides
- participation %high
- participation %low
- pauses
- perception %enhanced
- periodical
- person
- person %familiar
- personal
- personal %low
- personal contact
- personal contact %low
- personal identification
- personalisation
- personality %none
- phone
- phone conference call
- phone satellite
- physical
- physical contact
- physical distance

- physical distance %low
- physical location
- physical presence
- planning speech
- pose %relaxed
- pose %unrelaxed
- positive
- possible
- posture
- predictable
- presence flag
- pressure reduction
- previous contact %none
- previous experience
- previous friendship
- priority
- priority %highest
- probability %low
- process end
- process middle
- protocol
- proxemics
- proximity %auditory
- proximity %physical
- quality %high
- quality %poor
- quantity
- quantity %high
- quantity %low
- random %high
- random %low
- rational %low
- rational distinction
- rationalisation
- reaction
- real appearance
- real life
- real person
- real person %no
- real time
- realism
- realism %low
- recollecting
- reinforcement
- relationship
- relevance %high
- relevance %low
- relevance %none
- reliance
- reliance %low
- remoteness %reduced
- repetitive
- representation
- requirement
- requirements
- resemblance
- resemblance %lacking
- resolution %none
- respect %low
- response
- responsive %low
- rl behaviour
- rl meeting
- robotic
- roleplay
- safety
- scenario
- screen
- seeing
- seeking feedback
- self
- separate entity

- setup
- shifty
- signals
- significance %high
- significance %low
- similarity scale
- simulated gaze
- skepticism
- smalltalk
- social
- social banter
- social etiquette
- social norms
- solution %impossible
- something missing
- spatial
- speech
- subtlety
- surroundings
- symmetry
- synch %poor
- task
- task completion time %greater
- task focus
- task performance
- technical consideration
- technology
- telephone
- television
- tension
- theoretical rejection
- thoughts %unknown
- time
- time lag
- time limitation
- time pressure
- time pressure %low
- tracking %none
- trust
- trust %beginning
- trust %enhanced
- trust %low
- trust %undermined
- turntaking
- turntaking disruption
- two way
- uncertainty
- understand objectives
- uniformity
- unnecessary
- user
- utility %low
- verbal content
- verbal message
- videoconferencing
- visual
- visual anonymity
- visual appearance
- visual bonding %none
- visual clues
- visual contact %none
- visual discrepancy
- visual focus
- visual focus %low
- visual focus %none
- visual focus%none
- visual identity
- visual image
- visual information
- visual realism
- vocal focus
- vocal irruption

- voice
- voice %timing
- voice quality
- voice tone
- while speaking
- working %uncertain

F.3 Process of Visual Abstraction

During the second phase of axial coding, the individual network diagrams were consolidated into a large A1-size global diagram, of which a detail is shown in Figure F.1. When this diagram became saturated, it became necessary to clarify the interconnections between concepts; individual diagrams were created for each of the categories that had emerged as the main 'attractors'.

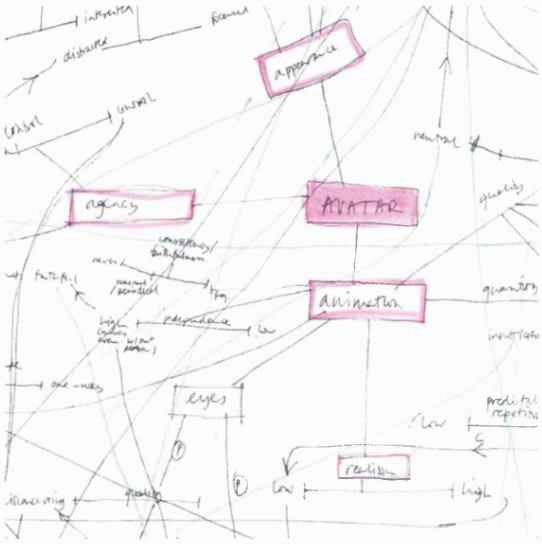


Figure F.1: Detail of the first global diagram consolidating the individual conceptual network diagrams

As mentioned in Chapter 6, a process of visual abstraction was developed to facilitate the final phase of selective coding. A colour key was created whereby each key category was represented by a colour (Figure F.4). A copy was made of the diagrams for each key category; by colouring the categories appearing on each diagram it became possible to identity interrelationships at a glance (see Figures F.2 and F.3). Some colours frequently appeared in close proximity, while others did not. Some appeared to be closely interconnected with the avatar subcategories, whereas others were more removed. For example, 'co-location' did not have direct connections to either the avatar's 'appearance ' or 'behaviour' categories, whereas 'personal contact' did. Most importantly, almost all of the categories appeared to connect either directly or indirectly to 'visual focus', recorded in the memos as being the probable core category.

Next, the categories were reduced to a coloured dot to abstract away from the verbal data entirely. At this stage the goal was to explore the main 'arteries' of traffic connecting the various key categories. The subcategories and properties were subsumed into these coloured points. On

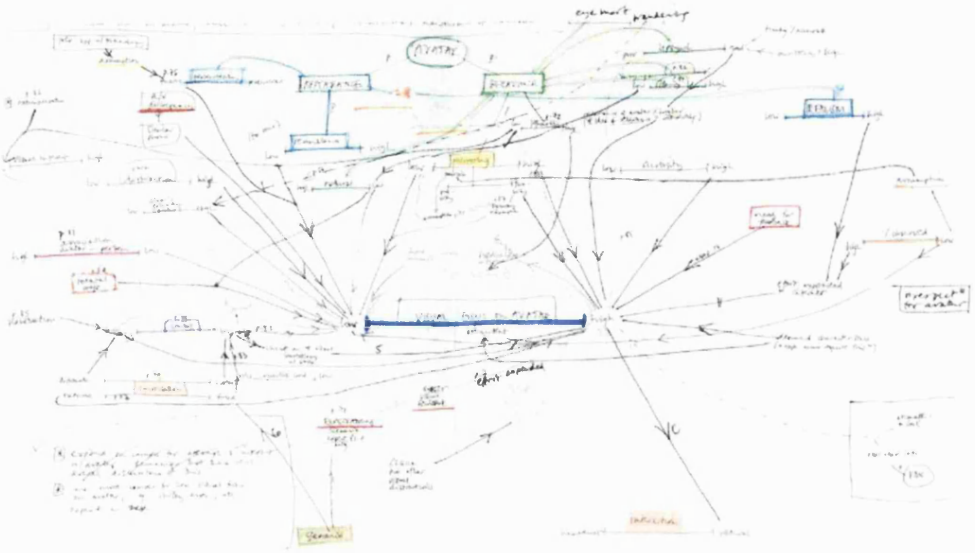


Figure F.2: Diagram focusing on the 'visual focus' category

Figure F.3: Diagram focusing on the 'trust' category

a new diagram, these coloured dots were linked according to the causal connections expressed in the data (Figure F.5). This diagram suggested that the 'visual focus' category appeared to be linked to most if not all other categories, either directly or indirectly. Another diagram was produced, where each coloured dot representing a key category was treated as the centre of its own mini-diagram; these mini-diagrams were produced on the same page for rapid comparison (Figure F.6).

Figure F.4: Colour key

Figure F.5: Diagram exploring the higher-level connections between key categories, now abstracted to coloured dots

Figure F.6: Mini-diagrams illustrating the higher-level connections for each key category

The overall picture confirmed that the dark blue dot representing 'visual focus' had the highest occurrence as well as the highest number of other categories connected to it (see Figure F.6, lower right). It also clarified that some categories appeared to be primarily causes, while others were primarily outcomes. Importantly for the purposes of this analysis, it also instantly underlined which categories were linked with the avatar categories and which were not. Two final diagrams were produced, this time breaking down stages of a process model. They ordered the categories according to whether they were primarily causes or outcomes (Figure F.7).

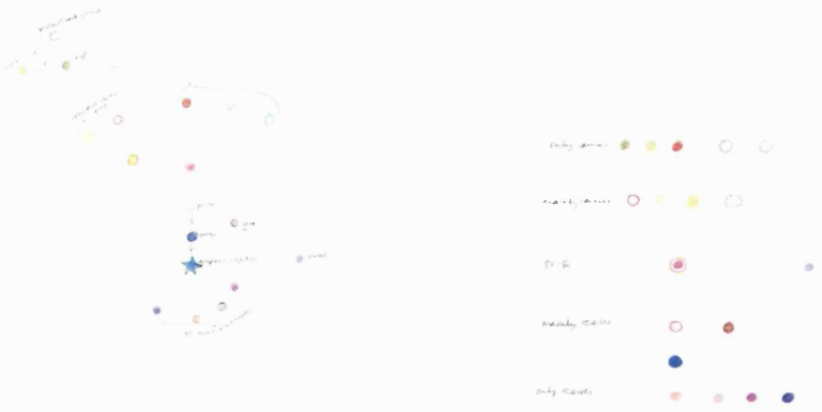


Figure F.7: Diagrams summarising the location of key categories in the process model

This process of replacing verbal data with colour ultimately helped to abstract away from the content of the categories to their interconnections, enabling the identification of a process model with a clear core category.

Appendix G

Additional images from the experiment on responsiveness

This appendix contains images of the psychophysiological monitoring equipment used in the experiment on responsiveness (reported in Chapter 4).

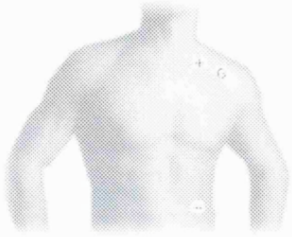


Figure G.1: Participants were asked to attach the EKG sensors as illustrated, with the positive and ground leads on the left collarbone and the negative lead on the lower left rib. This arrangement is recommended by Meehan [Mee01] to maximise the amplitude of the signal while minimising artifacts.

Figure G.2: The bipolar EDA sensors were attached as illustrated to the index and middle fingers of the participant's non-dominant hand.

Figure G.3: The psychophysiological monitoring equipment was placed inside a wearable pack for the participant's comfort. This also ensured that the delicate optical cable remained protected, minimising signal failures.

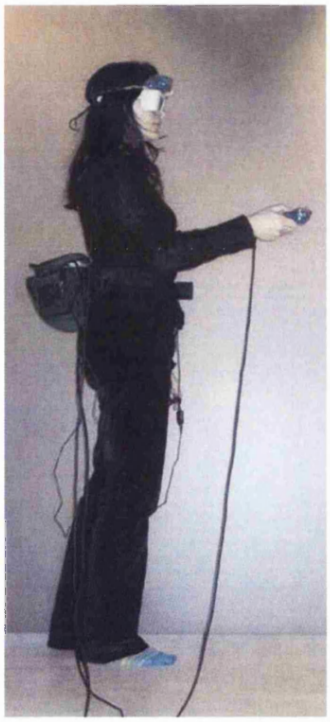


Figure G.4: Participants wore stereoscopic goggles and a head-tracking device, and held the 3D mouse in their dominant hand. The psychophysiological monitoring equipment was secured in a pack strapped to their waist.

Appendix H

Acronyms

BVP	Blood volume pulse
CVE	Collaborative virtual environment
EDA	Electrodermal activity
FOV	Field of view
GSR	Galvanic skin response
HMD	Head-mounted display
HR	Heart rate
IVE	Immersive virtual environment
PIP	Picture-in-picture approach (for two-party and multiparty videoconferences)
POV	Point of view
SAD	Social Avoidance and Distress questionnaire
SVE	Shared virtual environment
VE	Virtual environment
VMC	Video-mediated communication

Bibliography

[AB98] N. Allbeck and N. Badler. Avatars à la Snow Crash. In *Proceedings of Computer Animation*, pages 19–24, Philadelphia, PA, 1998. IEEE Computer Society.

[AC76] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge, 1976.

[Act03] Activeworlds. `http://www.activeworlds.com/`, July 2003.

[AD65] M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, 28:289–304, 1965.

[AI72] M. Argyle and R. Ingham. Mutual gaze and proximity. *Semiotica*, 6:32–49, 1972.

[AIAM73] M. Argyle, R. Ingham, F. Alkema, and M. McCallin. The different functions of gaze. *Semiotica*, 7:10–32, 1973.

[ALC74] M. Argyle, L. Lefebvre, and M. Cook. The meaning of five patterns of gaze. *Eur. J. Soc. Psychol.*, 4(2):125–136, 1974.

[And80] J. Andreassi. *Psychophysiology: Human behavior and physiological response*. Oxford University Press, Oxford, 1980.

[Arg88] M. Argyle. *Bodily Communication*. Methuen & Co., London, 2nd edition, 1988.

[ATL03] ATLAS.ti: The knowledge workbench. `http://www.atlasti.de/`, July 2003.

[AV02] A.L. Albuquerque and L. Velho. Togetherness through virtual worlds: How real can be that presence? In *Fifth Annual International Workshop: Presence 2002*, pages 435–447, Porto, Portugal, 2002.

[BBBL01] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virutal Environments*, 10(6):583–598, 2001.

[BBF+94] S. Benford, J. Bowers, L.E. Fahlén, J. Mariani, and T. Rodden. Supporting cooperative work in virtual environments. *The Computer Journal*, 37(8):653–668, 1994.

[BBF+95] S. Benford, J. Bowers, L.E. Fahlén, C. Greenhalgh, and D. Snowdon. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 242–249, Denver, CO, 1995.

[BBHS01] F. Biocca, J. Burgoon, C. Harms, and M. Stoner. Criteria and scope conditions for a theory and measure of social presence. In *Fourth Annual International Workshop: Presence 2001*, Philadelphia (PA), USA, May 2001. Available: `http://astro.temple.edu/~lombard/P2001/Biocca1.pdf`.

[BBRG96] S. Benford, C. Brown, G. Reynard, and C. Greenhalgh. Shared spaces: transportation, artificiality, and spatiality. In *Proceedings of the ACM conference on Computer supported cooperative work*, pages 77–85, Cambrdige, MA, 1996.

[BC02] F. Barrientos and J.F. Canny. Cursive: controlling expressive avatar gesture using pen gesture. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments*, pages 113–119, Bonn, Germany, 2002. ACM Press.

[BDR00] S. Benford, P. Dourish, and T. Rodden. Introduction to the special issue on human-computer interaction and collaborative virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(4):439–441, 2000.

[BGRP01] S. Benford, C. Greenhalgh, T. Rodden, and J. Pycock. Collaborative virtual environments: To what extent is cyberspace really a space? *Communications of the ACM*, 44(7):79–85, July 2001.

[BH02] F. Biocca and C. Harms. Defining and measuring social presence: Contribution to the networked minds theory of social presence. In *Fifth Annual International Workshop: Presence 2002*, pages 7–36, Porto, Portugal, 2002.

[BHB02] F. Biocca, C. Harms, and J.K. Burgoon. Criteria for a theory and measure of social presence. *Presence: teleoperators and virtual environments*, 2002. (Under review). Available: `http://www.mindlab.org/networkedminds/`.

[BHSS00] C. Basdogan, C.H. Ho, M.A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction*, 7(4):443–460, December 2000.

[Bio97] F. Biocca. The cyborg's dilemma: Embodiment in virtual environments. *Journal of CMC*, 3(2):12–26, 1997.

[BK02] G. Bente and N. Kramer. Virtual gestures: Analysing social presence effects of computer-mediated and computer-generated nonverbal behaviour. In *Fifth Annual International Workshop: Presence 2002*, pages 233–244, Porto, Portugal, 2002.

[Bla02] J. Blascovich. Social influence within immersive virtual environments. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 127–145. Springer-Verlag, London, 2002.

[BLB+01] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13:146–149, 2001.

[BM02] B. Becker and G. Mark. Social conventions in computer-mediated communication: A comparison of three online shared virtual environments. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 19–39. Springer-Verlag, London, 2002.

[BPO96] J. Bowers, J. Pycock, and J. O'Brien. Talk and embodiment in collaborative virtual enviroments. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 58–65, Vancouver, British Columbia, Canada, 1996.

[Bul83] P. Bull. *Body movement and interpersonal communication.* Wiley, Chichester, 1983.

[Bux92] W. Buxton. Telepresence: Integrating shared task and person spaces. In *Proceedings of the conference on graphics interface*, pages 123–129, Vancouver, BC, Canada, 1992. Morgan Kaufmann Publishers Inc.

[BW93] W. Barfield and S. Weghorst. The sense of presence within virtual environments: A conceptual framework. In G. Salvendy and M. Smith, editors, *Human Com-*

puter Interaction: Software and Hardware Interfaces, pages 699–704. Elsevier, Amsterdam, 1993.

[BY98] V. Bruce and A. Young. *In the Eye of the Beholder: The Science of Face Perception*. Oxford University Press, Oxford, 1998.

[CBB+99] J. Cassell, T. Bickmore, M. Billinghurst, K. Campbell, H. Chang, H. Vilhjálmsson, and H. Yan. Embodiment in conversational agents: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 520–527, Pittsburgh, PA, 1999.

[CCD00] A. Colburn, M. Cohen, and S. Drucker. The role of eye gaze in avatar-mediated conversational interfaces. Technical Report MSR-TR-2000-81, Microsoft Research, 2000.

[CFS02] L. Cheng, S. Farhnam, and L. Stone. Lessons learned: Building and deploying shared virtual environments. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work. Springer-Verlag, London, 2002.

[Cha95] K. Charmaz. Grounded theory. In J.A. Smith, R. Harre, and L. Langenhove, editors, *Rethinking Methods in Psychology*, pages 27–49. Sage Publications, London, 1995.

[CK98] C. Cieurzo and M.A. Keitel. Ethics in qualitative research. In M. Kopala and L.A. Suzuki, editors, *Using Qualitative Methods in Psychology*, pages 63–75. Sage Publications, Thousand Oaks, CA, 1998.

[CNSD93] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142, 1993.

[Cor96] R. Cornelius. *The Science of Emotion*. Prentice-Hall, Upper Sadle River, NJ, 1996.

[CT90] J. Cacioppo and L. Tassinary. *Principles of psychophysiology : Physical, social, and inferential elements*. Cambridge University Press, Cambridge, 1990.

[CW00] E. Cuddihy and D. Walters. Embodied interaction in social virtual environments. In *Proceedings of the third international conference on Collaborative virtual environments*, pages 181–188, San Francisco, CA, September 2000.

[Dey99] I. Dey. *Grounding Grounded Theory*. Academic Press, London, 1999.

[DGI00] P. De Greef and W.A. IJsselsteijn. Social presence in the photoshare tele-application. In *Proceedings of the 3rd International Workshop on Presence*, Delft, The Netherlands, March 2000. Available: `http://www.presence-research.org/papers/socpres.html`.

[DKF02] C. Dillon, E. Keogh, and J. Freeman. It's been emotional: Affect, physiology and presence. In *Fifth Annual International Workshop: Presence 2002*, pages 223–232, Porto, Portugal, 2002.

[DKU98] J.V. Draper, D.B. Kaber, and J.M. Usher. Telepresence. *Human Factors*, 40(3):354–375, 1998.

[DS00] N. Durlach and M. Slater. Presence in shared virtual environments and virtual togetherness. *Presence: Teleoperators and Virtual Environments*, 9(2):214–217, 2000.

[EF69] P. Ekman and W.V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, 1:49–98, 1969.

[EFE72] P. Ekman, W.V. Friesen, and P. Ellsworth. *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Pergamon Press, New York, 1972.

[EGR91] C.A. Ellis, S.J. Gibbs, and G.L. Rein. Groupware: Some experiences and issues. *Communications of the ACM*, 34(1):38–58, January 1991.

[ES00] S.M. Ehrlich and D. Schiano. Communicating facial affect: It's not the realism, it's the motion. In ACM Press, editor, *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 252–253, 2000.

[Eye03] Eyematic. `http://www.activeworlds.com/`, July 2003.

[FAM+00] J. Freeman, S.E. Avons, R. Meddis, D.E. Pearson, and W. IJsselsteijn. Using behavioural realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators and Virtual Environments*, 9(2):149–164, 2000.

[FAPI99] J. Freeman, S.E. Avons, D.E. Pearson, and W. IJsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments*, 8(1):1–13, February 1999.

[FGV+00] M. Fraser, I. Glover, I. Vaghi, S. Benford, C. Greenhalgh, J. Hindmarsh, and C. Heath. Revealing the realities of collaborative virtual reality. In *Proceedings of the Third International Conference on Collaborative Virtual Environments*, pages 29–37, San Francisco, CA, September 2000.

[FLI01] J. Freeman, J. Lessiter, and W. IJsselsteijn. An introduction to presence: A sense of being there in the mediated environment. *The Psychologist, British Psychological Society*, April 2001.

[For85] J.P. Forgas. *Interpersonal behaviour: The psychology of interpersonal interaction.* Pergamon Press, Oxford, 1985.

[FPW98] L. Fuchs, S. Poltrock, and R. Wojcik. Business value of 3D virtual environments. *Fuchs, L. and Poltrock, S. and Wojcik, R.*, 19(3):25–29, December 1998.

[FSO+01] A. Fukayama, M. Sawaki, T. Ohno, H. Murase, N. Hagita, and N. Mukawa. Expressing personality of interface agents by gaze. In *Proceedings of INTERACT Conference on Human-computer Interaction*, pages 793–794, Tokyo, Japan, 2001.

[FSS+01] E. Frécon, G. Smith, A. Steed, M. Stenius, and O. Stahl. An overview of the COVEN platform. *Presence: Teleoperators & Virtual Environments*, 10:109–127, 2001.

[FTM+02] A. Fukayama, O. Takehiko, N. Mukawa, M. Sawaki, and N. Hagita. Messages embedded in gaze of interface agents: Impression management with agent's gaze. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 41–48, Minneapolis, MN, 2002. ACM Press.

[Gen03] Generalised Linear Interactive Model (GLIM). `http://www.nag.co.uk/stats/GDGE_soft.asp`, July 2003.

[Gid84] A. Giddens. *The constitution of society: Outline of the theory of structuration.* Polity, Cambridge, 1984.

[GL01] G. Goebbels and V. Lalioti. Co-presence and co-working in distributed collaborative virtual environments. In *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation*, pages 109 – 114, Cape Town, South Africa, 2001.

[Gof63] E. Goffman. *Behaviour in public places: Notes on the social organization of gatherings*. Collier-Macmillan Ltd., London, 1963.

[Gof71] E. Goffman. *The presentation of self in everyday life*. Penguin, Hammondsworth, 1971.

[Gol96] D. Goleman. *Emotional Intelligence*. Bloomsbury Publishing Plc., London, 1996.

[GP63] J.J. Gibson and A. Pick. The perception of another person's looking behaviour. *American Journal of Psychology*, 46:386–394, 1963.

[GSBS01] M. Garau, M. Slater, S. Bee, and A. Sasse. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seattle, WA, 2001.

[GSPR03] M. Garau, M. Slater, D-P. Pertaub, and S. Razzaque. The response of people to avatars in an immersive virtual environment. In *Presence: Teleoperators and Virtual Environments. Special Issue on Acting and Perceiving in Virtual Enviroments*, 2003. (In press).

[GSV+03] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536, Ft. Lauderdale, FL, April 2003.

[Hal66] E.T. Hall. *The Hidden Dimension: Man's Use of Space in Public and Private*. Doubleday, New York, 1966.

[HBAM97] S. Harrison, S. Bly, S. Anderson, and S. Minneman. The media space. In K.E. Finn, A. Sellen, and S.B. Wilbur, editors, *Video-Mediated Communication*, pages 273–300. Laerence Erblaum Associates, Mahwah, NJ, 1997.

[HD92] R.M. Held and N.I. Durlach. Telepresence. *Presence: Teleoperators and Virtual Environments*, 1(1):109–112, 1992.

[HD96] S. Harrison and P. Dourish. Re-placeing space: The roles of place and space in collaborative systems. In *Computer-Supported Cooperative Work*, pages 67–76, Cambridge, MA, 1996.

[Hee92] C. Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1:262–271, 1992.

[Hee01] C. Heeter. Reflections on real presence by a virtual person. Keynote speech delivered at the Fourth International Workshop on Presence, June 2001. Available: http://astro.temple.edu/ lombard/P2001/Heeter.pdf.

[Hen96] K.L Henwood. Qualitative inquiry: Perspectives, methods and psychology. In J.T.E. Richardson, editor, *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*, pages 25–40. The British Psychological Society, Leicester, 1996.

[HFH+98] J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Fragmented interaction: establishing mutual orientation in virtual environments. In *Proceedings of the ACM conference on Computer supported cooperative work*, pages 217–226, Seattle, WA, November 1998.

[HRK+95] L. F. Hodges, B. O. Rothbaum, R. Kooper, D. Opdyke, T. Meyer, M. North, and J. J. de Graff. Virtual environments for treating the fear of heights. *IEEE Computer*, 28(7):27–34, 1995.

[Hum03] Humanoid Animation Working Group (H-Anim). `http://www.hanim.org`, July 2003.

[IDRFA00] W. IJsselsteijn, H. De Ridder, J. Freeman, and S.E. Avons. Presence: Concept, determinants and measurement. In *Proceedings of SPIE Human Vision and Electronic Imaging*, pages 23–28, San Jose, CA, January 2000.

[IFdR01] W. IJsselsteijn, J. Freeman, and H. de Ridder. Presence: Where are we? *Cyberpsychology and Behavior*, 4(2):179–182, 2001.

[IKG93] H. Ishii, M. Kobayashi, and J. Grudin. Integration of interpersonal space and shared workspace: Clearboard design and experiments. *ACM Transactions on Information Systems (TOIS)*, 11(4):349–375, October 1993.

[Iza71] C.E. Izard. *The Face of Emotion*. Appleton-Century-Crofts, New York, NY, 1971.

[Iza97] C.E. Izard. Emotions and facial expressions: A perspective from differential emotions theory. In J.A. Russell and J-M. Fernandez-Dols, editors, *The Psychology of Facial Expression*. Cambridge University Press, Cambridge, 1997.

[Jak02] M. Jakobsson. Rest in peace, bill the bot: Death and life in virtual worlds. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 63–76. Springer-Verlag, London, 2002.

[KB97] T. Kim and F. Biocca. Telepresence via television: Two dimensions of telepresence may have difference connections to memory and persuasion. *Journal of Computer-Mediated Communication*, 3(2):Available: `http://www.ascusc.org/jcmc/vol3/issue2/kim.html`, 1997.

[KDKO02] R. Kaushik, P. David, S. Kline, and D.J. Oaks. Social presence effects: A study of CMC vs. FtF in a collaborative fiction project. In *Fifth Annual International Workshop: Presence 2002*, pages 395–420, Porto, Portugal, 2002.

[Ken67] A. Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63, 1967.

[KI97] M. Katsikitis and J.M. Innes. Encoding and decoding of facial expression. *The Journal of General Psychology*, 124(4):357–370, 1997.

[KKER98] K. Kauppinen, A. Kivimäki, T. Era, and M. Robinson. Producing identity in collaborative virtual environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)*, 1998.

[KLCH00] D.H. Krikorian, J.S. Lee, T.M. Chock, and C. Harms. Isn't that spatial? Distance and communication in a 2d virtual environment. *Journal of Computer-Mediated Communication (JCMC)*, 5(4), 2000. Available: `http://www.ascusc.org/jcmc/vol5/issue4/krikorian.html`.

[KS02] P. Kauff and O. Schreer. An immersive 3d video-conferencing system using shared virtual team user environments. In *Proceedings of the 4th international conference on Collaborative virtual environments*, pages 105–112, Bonn, Germany, 2002.

[LBB99] J. Loomis, J. Blascovich, and A. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior Reserach Methods, Instruments and Computers*, 31(4):557–564, 1999.

[LBB02] S.H. Lee, J.B. Badler, and N.I. Badler. Eyes alive. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 637–644, San Antonio, TX, 2002.

[LD97] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997. Available: `http://www.ascusc.org/jcmc/vol3/issue2/lombard.html`.

[LGPW98] C. Lee, S. Ghyme, C. Park, and K. Wohn. The control of avatar motion using hand gesture. In *Proceedings of the ACM symposium on Virtual reality software and technology 1998*, pages 59–65, Taipei, Taiwan, 1998. ACM Press.

[LN01] K.M. Lee and C. Nass. Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. In *4th Anuual International Workshop on Presence*, Philadelphia, PA, 2001.

[Lom00] M. Lombard. Resources for the study of presence: Presence explication, July 2000. Retrieved: `http://nimbus.temple.edu/~mlombard/Presence/explicat.htm`.

[Lud96] P. Ludlow. *High noon on the electronic frontier: conceptual issues in cyberspace.* MIT Press, Cambridge, MA, 1996.

[Mar03] Marilyn avatar, Miralab. `http://www.miralab.unige.ch/subpages/marilyn/flash/index.htm`, July 2003.

[Mat99] The Matrix. 16 mm, 136 min. Burbank: Warner Bros., 1999.

[MC98] K. Mania and A. Chalmers. A classification for user embodiment in collaborative virtual environments. In *Proceedings of the 4th International Conference on Virtual Systems and Multimedia (VSMM)*, volume 1, pages 177–182, 1998.

[Mee01] M.J. Meehan. *Physiological Reaction as an Objective Measure of Presence in Virtual Environments.* Phd thesis, University of North Carolina at Chapel Hill, 2001.

[Mer98] E. Merrick. An exploration of quality in qualitative research: Are reliability and validity relevant? In M. Kopala and L.A. Suzuki, editors, *Using Qualitative Methods in Psychology*, pages 25–36. Sage Publications, Thousand Oaks, CA, 1998.

[MF90] C. Morningstar and R. Farmer. The lessons of Lucasfilm's Habitat. In Michael Benedikt, editor, *Cyberspace: First Steps*. The MIT Press, Cambridge, MA, 1990.

[MIWB01] M.J. Meehan, B. Insko, M. Whitton, and F. Brooks. Objective measures of presence in virtual environments. In *Presence 2001 4th International Workshop*, Philadelphia, USA, 2001.

[MN83] P. McCullagh and J.A. Nelder. *Generalised Linear Models*. Chapman and Hall, 1983.

[MSP+03] M.L. McLaughlin, G. Sukhatme, W. Peng, W. Zhu, and J. Parks. Performance and co-presence in heterogeneous haptic collaboration. In *IEEE Virtual Reality Haptics Symposium*, Los Angeles, CA, 2003.

[MVS+02] J. Mortensen, V. Vinayagamoorthy, M. Slater, A. Steed, B. Lok, and M. Whitton. Collaboration in tele-immersive environments. In W. Stürzlinger and S. Müller, editors, *Proceedings of the Eighth Eurographics Workshop on Virtual Environments (EGVE-02)*, pages 93–102, New York, 30–31 2002. ACM Press.

[NB01] K. Nowak and F. Biocca. Understanding the influence of agency and anthropomorphism on copresence, social presence and physical presence with virtual humans. In *4th Annual International Workshop on Presence*, Philadelphia, USA, 2001.

[NHSA02] A. Nilsson, I. Heldal, R. Schroeder, and A.S. Axelsson. The long-term uses of shared virtual environments: An exploratory study. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 112–126. Springer-Verlag, London, 2002.

[Now01] K. Nowak. Defining and differentiating copresence, social presence and presence as transportation. In *Fourth Annual International Workshop: Presence 2001*, Philadelphia (PA), USA, 2001. Available: `http://astro.temple.edu/~lombard/P2001/papers.htm`.

[OMIM94] K.I. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita. Multiparty videoconferencing at virtual social distance: Majic design. In *Proceedings of the 1994 ACM Conference on Computer-Supported Cooperative Work*, pages 385–393, Chapel Hill, NC, 1994.

[Onl03] Onlive!Traveler. `http://www.digitalspace.com/traveler/`, July 2003.

[Pal03] The Palace. `http://www.thepalace.com/`, July 2003.

[PG98] J.G. Ponterotto and I. Grieger. Merging qualitative and quantitative perspectives in a research identity. In M. Kopala and L.A. Suzuki, editors, *Using Qualitative Methods in Psychology*, pages 49–62. Sage Publications, Thousand Oaks, CA, 1998.

[PH96] N. Pidgeon and K. Henwood. Grounded theory: Practical implementation. In J.T.E. Richardson, editor, *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*, pages 86–101. The British Psychological Society, Leicester, 1996.

[Pic97] R. Picard. *Affective Computing*. MIT Press, Cambridge, MA, 1997.

[Pid96] N. Pidgeon. Grounded theory: Theoretical background. In J.T.E. Richardson, editor, *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*, pages 75–85. The British Psychological Society, Leicester, 1996.

[PSB01a] D-P. Pertaub, M. Slater, and C. Barker. An experiment on fear of public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11(1):68–78, 2001.

[PSB01b] D-P. Pertaub, Mel Slater, and C. Barker. An experiment on fear of public speaking in virtual reality. In Westwood, editor, *Medicine Meets Virtual Reality*, pages 372–378, 2001.

[RB02] D.M. Romano and P. Brna. Collaboration, presence and performance in virtual learning environments: Can collaboration be used to measure shared presence? In *Fifth Annual International Workshop: Presence 2002*, pages 129–138, Porto, Portugal, 2002.

[RN96] B. Reeves and C. Nass. *The media equation: How people treat computers television and new media like real people and places.* Cambridge University Press, Cambridge U.K., 1996.

[Sal99] E-L. Sallnäs. Presence in multimodal interfaces. In *2nd International Workshop on Presence*, Colchester, UK, April 1999.

[Sal02] E-L. Sallnäs. Collaboration in multi-modal virtual worlds: Comparing touch, text, voice and video. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 172–187. Springer-Verlag, London, 2002.

[SB02] K. Swinth and J. Blascovich. Perceving and responding to others: Human-human and human-computer social interaction in collaborative virtual environments. In *Fifth Annual International Workshop: Presence 2002*, pages 310–340, Porto, Portugal, 2002.

[SC98] A. Strauss and J. Corbin. *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Thousand Oaks, CA, 2nd edition, 1998.

[Sch96] R. Schroeder. *Possible worlds: The social dynamic of virtual reality technologies*. Westview Press, Boulder, 1996.

[Sch02a] R. Schroeder. Copresence and interaction in virtual environments: An overview of the range of issues. In *Fifth Annual International Workshop: Presence 2002*, pages 274–295, Porto, Portugal, 2002.

[Sch02b] R. Schroeder. An overview of the contributions. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages ix–xiv. Springer-Verlag, London, 2002.

[Sch02c] R. Schroeder. Social interaction in virtual environments: Key issues, common themes, and a framework for research. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 1–18. Springer-Verlag, London, 2002.

[Sci98] D. Sciarra. The role of the quantitative researcher. In M. Kopala and L.A. Suzuki, editors, *Using Qualitative Methods in Psychology*, pages 37–48. Sage Publications, Thousand Oaks, CA, 1998.

[SEKS00] D. Schiano, S.M. Ehrlich, R. Krisnawan, and K. Sheridan. Face to interface: Facial affect in (hu)man and machine. In ACM Press, editor, *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 193–200, 2000.

[Sel95] A. Sellen. Remote conversations: The effect of mediating talk with technology. *Human-Computer Interaction*, 10(4):401–444, 1995.

[SFD02] M. Smith, S. Farnham, and S. Drucker. The social life of small graphical chat spaces. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 205–220. Springer-Verlag, London, 2002.

[She92a] T.B. Sheridan. Defining our terms. *Presence: Teleoperators and Virtual Environments*, 1(2):272–, 1992.

[She92b] T.B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1:120–125, 1992.

[She96] T.B. Sheridan. Further musings on the psychophysics of presence. *Presence: Teleoperators and Virtual Environments*, 5(2):241–246, 1996.

[SHS+00] M. Slater, J. Howell, A. Steed, D-P. Pertaub, M. Garau, and S. Springel. Acting in virtual reality. In *Proceedings of the Third International Conference on Collaborative Virtual Environments*, pages 103–110, San Francisco, CA, 2000.

[Sla99a] M. Slater. Measuring presence: A response to the witmer and singer questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5):560–565, 1999.

[Sla99b] M. Slater. Real people meeting virtually real people: A review of some experiments in shared virtual enviroments. *BT Technology Journal*, 17(1):120–127, 1999.

[Sla04] M. Slater. How colourful was your day? Why questionnaires cannot assess presence in virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 2004. (In press).

[SM94] S. Straus and J.E. McGrath. Does the medium matter? The interaction of task and technology on group performance and member reactions. *Journal of Applied Psychology*, 79:87–97, 1994.

[Smi95] J.A. Smith. Semi-structured interviewing and qualitative analysis. In J.A. Smith, R. Harre, and L. Langenhove, editors, *Rethinking Methods in Psychology*, pages 9–26. Sage Publications, London, 1995.

[SS72] A.E. Scheflen and A. Scheflen. *Body language and social order: Communication and behavioral control.* Prentice-Hall, Englewood Cliffs, 1972.

[SS00] M. Slater and A. Steed. A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, 9(5):413–434, 2000.

[SS02] M. Slater and A. Steed. Meeting people virtually: Experiments in shared virtual environments. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 146–171. Springer-Verlag, London, 2002.

[SS03] O. Schreer and P. Sheppard. Virtue: The step towards immersive telepresence in virtual video-conference systems. Available: `http://www.virtue.eu.com/PublicArea2001/papers/eBusiness'00.pdf`, July 2003.

[SSA+01] R. Schroeder, A. Steed, A-S. Axelsson, I. Heldal, Å. Abelin, J. Wideström, A. Nilsson, and M. Slater. Collaborating in networked immersive spaces: As good as being there together? *Computers & Graphics*, 25:781–788, 2001.

[SSC01] M. Slater, A. Steed, and Y. Chrysanthou. *Computer Graphics and Virtual Environments: From Realism to Real-Time.* Addison Wesley Publishers, Harlow, England, 2001.

[SSMM98] M. Slater, A. Steed, J. Mc.Carthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors*, 40(3):469–477, 1998.

[SSS+99] A. Steed, M. Slater, A. Sadagic, A. Bullock, and J. Tromp. Leadership and collaboration in shared virtual environments. In *IEEE Virtual Reality*, pages 58–63, Houston, 1999.

[SSU93] M. Slater, A. Steed, and M. Usoh. The virtual treadmill: A naturalistic method for navigation in immersive virtual environments. In *First Eurographics Workshop on Virtual Environments*, pages 71–83, 1993.

[SSUS00] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small group behaviour in a virtual and real environment: A comparative study. *Presence: Teleoperators and Virtual Environments*, 9(1):37–51, 2000.

[Ste92] N. Stephenson. *Snow Crash.* ROC, London, 1992.

[Sti93] W.B. Stiles. Quality control in qualitative research. *Clinical Psychology Review*, 13:593–618, 1993.

[SU93] M. Slater and M. Usoh. Presence in immersive virtual enviroments. In *IEEE Virtual Reality Annual International Symposium (VRAIS)*, pages 90–96, Seattle, WA, 1993.

[SU95] M. Slater and M. Usoh. Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer Human Interaction (TOCHI): Special Issue on Virtual Reality Software and Technology*, 2(3):201–219, 1995.

[Sul99] J. Suler. The psychology of avatars and graphical space in multimedia chat communities: Or how I learned to stop worrying and love my Palace props, 1999. Available: `http://www.rider.edu/~suler/psycyber/psyav.html`.

[SUS94] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3(2):130–144, 1994.

[SW97] M. Slater and S. Wilbur. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 1997.

[SWC76] J. Short, E. Williams, and B. Christie. *The Social Psychology of Telecommunications*. John Wiley & Sons, London, 1976.

[Tay02] T.L. Taylor. Living digitally: Embodiment in virtual worlds. In R. Schroeder, editor, *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*, Computer Supported Cooperative Work, pages 40–62. Springer-Verlag, London, 2002.

[TBS+98] J. Tromp, A. Bullock, A. Steed, A. Sadagic, M. Slater, and E. Frécon. Small group behaviour experiments in the COVEN project. *IEEE Computer Graphics and Applications*, 18(6):53–63, 1998.

[Tel03] Televirtual. `http://www.televirtual.com`, July 2003.

[Tho03] Thought Technology. `www.thoughttechnology.com/procomp.htm`, July 2003.

[TJ81] F. Thomas and O. Johnston. *Disney Animation: The Illusion of Life*. Abberville Press Publishers, New York, 1981.

[TR00] M.J. Taylor and S.R. Rowe. Gaze communication using semantically consistent spaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 400–407, The Hague, The Netherlands, 2000. ACM Press.

[TS97] J. Tromp and D. Snowdon. Virtual body language: Providing appropriate user interfaces in collaborative virtual environments. In ACM, editor, *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)*, pages 37–44, Lausanne, Switzerland, 1997.

[UCAS00] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, 2000.

[VC98] H. Vilhjálmsson and J. Cassell. Bodychat: Autonomous communicative behaviors in avatars. In *Proceedings of the 2nd Annual ACM International Conference on Autonomous Agents*, pages 269–276, Minneapolis, MN, 1998.

[VC99] H. Vilhjálmsson and J. Cassell. Fully embodied conversational avatars: Making communicative behaviours autonomous. *Autonomous Agents and Multi-agent Systems*, 2:45–64, 1999.

[Ver99] R. Vertegaal. The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. In ACM Press, editor, *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 294–301, Pittsburgh, USA, 1999.

[Wal98] H.G. Walbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.

[WAS+00] J. Wideström, A-S. Axelsson, R. Schroeder, A. Nisson, I. Heldal, and Å. Abelin. The collaborative cube puzzle: A comparison of virtual and real environments. In *Proceedings of the Third International Conference on Collaborative Virtual Environments*, pages 165–171, San Francisco, CA, 2000.

[WF69] D. Watson and R. Friend. Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, 33:448–457, 1969.

[Whi95] S. Whittaker. Rethinking video as a technology for interpersonal communications: Theory and design implications. *Int. J. Human-Computer Studies*, 1995(42):501–529, 1995.

[Wil01] C. Willig. *Introducing Qualitative Research in Psychology: Adventures in Theory and Method.* Open University Press, Buckingham, 2001.

[WO97] S. Whittaker and B. O'Connail. The role of vision in face-to-face and mediated communication. In K.E. Finn, A. Sellen, and S.B. Wilbur, editors, *Video-Mediated Communication*, pages 23–49. Lawrence Erblaum Associates, Mahwah, NJ, 1997.

[WS98] B.G. Witmer and M.J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[WS00] G. Wilson and M.A. Sasse. Do users always know what's good for them? Utilising physiological responses to assess media quality. In S. McDonald, Y. Waern, and G. Cockton, editors, *HCI 2000: People and Computers XIV-Usability or Else!*, pages 327–339, Sunderland, UK, September 2000. Springer.

[Zel92] D. Zeltzer. Autonomy, interaction, and presence. *Presence: Teleoperators and Virtual Environments*, 1:127–132, 1992.

[Zha01] S. Zhao. Toward a taxonomy of copresence. In *Fourth Annual International Workshop: Presence 2001*, Philadelphia, PA, 2001.